

UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS HUMANAS, LETRAS E ARTES
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

GUSTAVO NISHIDA

SOBRE TEORIAS DE PERCEPÇÃO DA FALA

CURITIBA
2012

GUSTAVO NISHIDA

SOBRE TEORIAS DE PERCEPÇÃO DA FALA

**Tese de doutorado
apresentada ao Programa de
Pós-Graduação em Letras da
Universidade Federal do
Paraná, como requisito
parcial para a obtenção de
título de Doutor em Letras,
área de concentração de
Estudos Linguísticos.**

**Orientadora: Prof^a. Dr^a.
Adelaide H.P. Silva**

CURITIBA

2012

Pra Helô.

AGRADECIMENTOS

Agradeço: a Helô; aos Nishidas, Kozievitchs e Ribeiros; aos professores doutores, que tanto contribuíram com este trabalho, Adelaide H.P. Silva, Denise Kluge, Gisela Collischonn, Izabel Seara, José Borges Neto, Kleber Candioto, Lígia Negri, Maria Lúcia Castro Gomes, Maximiliano Guimarães e Sandra Madureira; aos amigos Carlos Eduardo, Felipe, Juarez, Michele, Paulo Henrique, Raul e Renata; aos companheiros do Laboratório de Estudos Fônicos Flávio Medina, Jeniffer Albuquerque, Dr^a. Luciane Costa e Dr^a. Rita Tonocchi; ao Odair; ao Programa de Pós-Graduação em Letras da UFPR; à CAPES pela bolsa; e aos colegas da Unicentro de Irati.

*“You can beat us with wires
You can beat us with chains
You can run out your rules
But you know you can't outrun the history train
I 've seen a glorious day ...”
(Peace like a river – Paul Simon, 1972)*



PARECER

Defesa de tese do doutorando GUSTAVO NISHIDA para obtenção do título de **Doutor em Letras**.

As abaixo assinadas ADELAIDE HERCÍLIA PESCATORI SILVA, GISELA COLLISCHONN, IZABEL CHRISTINE SEARA, LÍGIA NEGRI e SANDRA MADUREIRA FONTES argüiram, nesta data, o candidato, o qual apresentou a tese:

“SOBRE TEORIAS DE PERCEPÇÃO DA FALA”

Procedida a argüição segundo o protocolo que foi aprovado pelo Colegiado do Curso, a Banca é de parecer que o candidato está apto ao título de **Doutor em Letras**, tendo merecido os conceitos abaixo:

Banca	Assinatura	APROVADO Não APROVADO
ADELAIDE H. PESCATORI SILVA		aprovado
GISELA COLLISCHONN		APROVADO
IZABEL CHRISTINE SEARA		APROVADO
LÍGIA NEGRI		Aprovado
SANDRA MADUREIRA FONTES		aprovado

Curitiba, 27 de novembro de 2012

Prof.ª Dr.ª Teresa Cristina Wachowicz
Vice-Coordenadora

Sumário

RESUMO.....	xii
ABSTRACT.....	xiv
RESUMEN.....	xvi
PREFÁCIO.....	xviii
INTRODUÇÃO.....	1
CAPÍTULO 1: A NATUREZA DA PERCEPÇÃO DA FALA.....	5
1 COMO AUDIÇÃO.....	7
1.1 Dois estágios nos processos perceptuais.....	8
1.2 Identificação por atributos secundários.....	19
1.3 Percepção dos sons por outros animais.....	30
1.4 Sumário da seção.....	32
2 COMO ARTICULAÇÃO.....	34
2.1 Como gestos pretendidos.....	34
2.1.1 Teoria Motora da Percepção da Fala.....	35
2.2 Como gestos reais.....	40
2.2.1 Teoria do Realismo Direto da Percepção da Fala.....	40
2.3 Sumário da seção.....	43
CAPÍTULO 2: DEMARCAÇÃO DO TERRENO DAS TEORIAS DE PERCEPÇÃO DA FALA.....	44
1 A PERCEPÇÃO DA FALA NO ESTRUTURALISMO.....	45
1.1 A Fonologia Estruturalista e o objeto saussureano.....	45
1.2 Mas afinal, o que é percepção?.....	48
1.2.1 Uma metodologia possível.....	48
1.2.2 A fisiologia da audição como linha promissora de pesquisa.....	51
1.3 Sumário da seção.....	52
2 A PERCEPÇÃO DA FALA NO GERATIVISMO.....	53
2.1 Uma tentativa de colocar tudo junto: a “Early Motor Theory”.....	53
2.2 Uma correção no meio do caminho.....	58
2.2.1 Pode ser a fala alfabética?.....	59
2.2.2 A reestruturação dos fonemas no som.....	60
2.2.3 O módulo da fala.....	64

2.2.4 Codificação e decodificação.....	65
2.3 A três passos da teoria.....	70
2.3.1 A Fonologia Gerativa: The Sound Pattern of English (Chomsky & Halle 1968).....	71
2.3.2 Ouvindo lábios e vendo vozes.....	74
2.3.3 A base da Teoria Motora: “A Modularidade da Mente” de Jerry Fodor (1983).....	79
2.3.3.1 Quatro tratamentos para a estrutura mental.....	79
2.3.3.2 A organização modular da mente.....	89
2.4 A Teoria Motora Revisada.....	92
2.4.1 Proposta.....	92
2.4.2 Evidências experimentais.....	95
2.4.2.1 Dados de coarticulação.....	96
2.4.2.2 Efeito McGurk.....	98
2.4.2.3 Percepção duplex.....	98
2.5 Sumário da seção.....	100
3 A PERCEPÇÃO DA FALA EM UM ARCABOUÇO DINÂMICO.....	101
3.1 Os primeiros achados.....	101
3.2 A base da TRD: Os sentidos como sistemas perceptuais (Gibson, 1966).....	114
3.2.1 O que são os sentidos?.....	115
3.2.2 Como obtemos estímulos?.....	117
3.2.3 Os sistemas perceptuais.....	123
3.2.4 A teoria da coleta de informação.....	128
3.3 Uma primeira formulação da Teoria do Realismo Direto.....	134
3.3.1 Proposta.....	134
3.3.2 Quatro barreiras para o Realismo Direto.....	138
3.3.2.1 Se os ouvintes recuperam a articulação, por que eles não sabem disso?.....	138
3.3.2.2 As unidades linguísticas não são literalmente articuladas.....	140
3.3.2.3 O sinal acústico não especifica os segmentos fonéticos.....	143
3.3.2.4 A percepção envolve processos top down e os percebedores realizam erros.....	146
3.3.3 Como a percepção guia a ação?.....	147
3.4 Fonologia Articulatória.....	150
3.5 Paridade: o conceito que faltava.....	157
3.6 Uma proposta unificada para a produção e a percepção da fala.....	162
3.7 Sumário da seção.....	166
CAPÍTULO 3: UMA POLÊMICA SOBRE A NATUREZA DOS PRIMITIVOS DE PERCEPÇÃO DA FALA.....	167
1 UMA POLÊMICA SOBRE A NATUREZA DOS PRIMITIVOS DE PERCEPÇÃO	

DA FALA.....	170
1.1 Entre o acústico e o articulatório: Ohala versus Fowler.....	171
1.2 Fazendo a vez de advogado do diabo ou como a TM responderia à polêmica....	177
1.3 O que é o acústico e o articulatório para cada uma das teorias de percepção da fala?.....	182
1.4 Sumário da seção: por que a polêmica não fecha?.....	189
CONSIDERAÇÕES FINAIS.....	194
REFERÊNCIAS BIBLIOGRÁFICAS.....	199

Lista de ilustrações

Ilustração 1: Resultados do teste de identificação absoluta (cf. Pisoni, 1973: p.256).....	11
Ilustração 2: Resultados para o procedimento (1) (cf. Pisoni, 1973: p. 256).....	12
Ilustração 3: Resultados para o procedimento (2) (cf. Pisoni, 1973: p.257).....	13
Ilustração 4: Proposta de Pisoni (1974: p.286) para os estágios de processamento da fala.....	16
Ilustração 5: Resultados para a tarefa de identificação (cf. Pisoni, 1974: p.2881). E possível verificar que os informantes dividiram o contínuo em duas categorias. A divisão ocorre aproximadamente aos 30ms de VOT. As linhas com os quadrados preenchidos mostram o comportamento de [ba] e os círculos vazados, [pa]. Já a linha pontilhada com os triângulos correspondem às latências de identificação de cada resposta.....	18
Ilustração 6: Espectrogramas dos estímulos utilizados no experimento de Cole & Scot (1974: p.360): consoantes sonoras.....	25
Ilustração 7: Espectrogramas dos estímulos utilizados no experimento de Cole & Scot (1974: p.361): consoantes surdas.....	25
Ilustração 8: Sistema de conversação entre A e B proposto por Sausurre (1969 : p.19).....	46
Ilustração 9: Padrões espectrográficos que produzem /di/, /da/ e /do/. As linhas pontilhadas mostram o locus de 1800cps1 para o /d/ (Liberman, 1957 apud Liberman, 1996: p.190).....	54
Ilustração 10: Padrões espectrográficos que produzem /d/ e /g/ em vários contextos vocálicos. As linhas pontilhadas indicam os loci das consoantes (Liberman, 1957 apud Liberman, 1996: p.194).	55
Ilustração 11: Padrões espectrográficos suficientes para sintetizar /d/ diante de /i/ e /u/ (Liberman e colegas, 1967: p.206).....	61
Ilustração 12: (A) Transições que se iniciam no locus de /d/. (B) Transições que apenas apontam para o locus de /d/ (Liberman e colegas, 1967: p.208).....	62
Ilustração 13: Representação esquemáticas dos estágios assumidos para a produção da fala. (Liberman e colegas, 1967: p.216).....	67
Ilustração 14: Organização dos componentes que integram a Fonologia proposta em SPE. Diagrama adaptado de Albano (2001: p.18).....	73
Ilustração 15: Tabela presente em McDonald & McGurk (1976: p.747) que apresenta as respostas obtidas em seu experimento.....	76
Ilustração 16: Tabela presente em McDonald & McGurk (1976: p.747) que apresenta as porcentagens das respostas dadas pelos informantes em cada uma das condições do experimento.....	77
Ilustração 17: Estímulos apresentados no experimento de percepção duplex (Liberman; Isenberg & Rakerd, 1981: p.134).....	99
Ilustração 18: Representação esquemática da coprodução de vogais e consoantes (Fowler, 1981:	

p.46).....	104
Ilustração 19: Esquemas de segmentação do sinal da fala (Fowler, 1984: p.360).....	108
Ilustração 20: Traçado cinefluorográfico do traço vocal enquanto produzia /husi/ (redesenhado de Carney & Moll, 1971): ____ é [u]; ----, [s]/ x-x-x-x, [i] (Fowler, 1986: p.10).....	141
Ilustração 21: (a) Espectrograma com a produção de “Santa Claus”. (b) Visualização esquemática da relação entre os segmentos acústicos e fonéticos (reimpresso de Cutting & Pisoni, 1978 e Fant & Lindblom, 1961 apud Fowler, 1986: p.12).....	144
Ilustração 22: Pauta gestual da palavra “bad” (baseada em Browman & Goldstein, 1992: p.158).	155
Ilustração 23: Pauta gestual da palavra “pad” (baseada em Browman & Goldstein, 1992: p.158).	156

Lista de tabelas

Tabela 1: Os sistemas perceptuais e seu detalhamento (Gibson, 1966: p.50).....	127
Tabela 2: Variáveis do trato e articuladores envolvidos na realização da tarefa especificada pelas variáveis.....	153

RESUMO

O objetivo geral deste trabalho é refletir sobre três diferentes maneiras de se conceber os primitivos adotados pelas teorias de perceptuais da fala. No capítulo 1 nos deteremos a apresentar que a percepção da fala pode ser considerada como sendo de base auditiva ou de base articulatória. Na primeira, argumenta-se a favor da existência de invariantes acústicos que promoveriam a categorização de unidades fônicas distintivas no sinal da fala. Na segunda, apresenta-se o gesto como primitivo perceptual, podendo ser abstrato (tal como sugere a Teoria Motora da Percepção da Fala – Liberman & Mattingly, 1985) ou real (Fowler, 1986 e seguintes).

Como objetivo específico, traçamos (no Capítulo 2) um breve percurso da constituição das teorias de percepção da fala. Nesse percurso histórico, vinculamos a proposição de primitivos perceptuais (tanto de base acústica quanto articulatória) com o advento de teorias fonológicas que adotavam novos primitivos de análise. Isto é, trata-se de verificar que a adoção de novos primitivos perceptuais está vinculada aos cortes epistemológicos sugeridos pelas teorias fonológicas.

Para exemplificar a nossa proposta, discutiremos (no Capítulo 3) o “debate” publicado em 1996 entre Carol Fowler e John Ohala sobre a natureza ou acústica ou articulatória da percepção da fala. O interesse sobre esse debate é que ele não se desenvolve, pois há apenas um texto de cada um dos participantes. Argumentamos que a “morte” da discussão indica que as teorias de percepção se assentam sobre terrenos epistemológicos distintos que impossibilitam o fechamento do debate.

Interessar-se pela natureza dos primitivos da percepção da fala (debatido por Fowler e Ohala) é importante porque até o Estruturalismo não se questionava qual era a sua natureza. Para os estruturalistas, a percepção era reduzida à audição, i.e., considerava apenas os dados que se ouvia como distintivo em um *corpus*. Ao que parece, o problema surge por acaso quando Liberman e colegas (cf. Liberman, 1957) tentavam encontrar o que havia de distintivo entre as categorias fonêmicas. Para isso, buscavam (com a ajuda do *Pattern Playback*: um aparelho que convertia espectrogramas desenhados à mão em som) quais eram as pistas acústicas que diferenciavam um [p] de um [d], por exemplo. No entanto, os resultados não apresentavam uma relação biunívoca entre o acústico e o articulatório, sugerindo que a articulação guiava a percepção; pois o que haveria de

invariante na fala era a articulação, uma vez que o sinal acústico apresentava diferenças dentro de uma mesma categoria devido à influência das vogais.

Nossa pesquisa mostrou que os primitivos perceptuais seguem as propostas de primitivos de análise das teorias fonológicas, a saber: no Estruturalismo, não havia uma teoria de percepção. No entanto, as pesquisas iniciais tentavam buscar os invariantes perceptuais acústicos influenciadas pela Fonologia Estruturalista de Jakobson, Fant & Halle (1952); no período gerativista, a Teoria Motora da Percepção da Fala (Liberman & Mattingly, 1985) propõe primitivos de base articulatória abstratos. Esse modelo ganha a sua versão revisada (e mais robusta) a partir da proposta de Chomsky & Halle (1968), dos achados referentes ao Efeito *McGurk* (McDonald & McGurk, 1976) e da teoria modular da mente de Fodor (1983); por fim, há a proposição da Teoria do Realismo Direto da Percepção da Fala (Fowler, 1986, e seguintes) influenciada pela teoria de percepção de Gibson (1966). Essa abordagem converge para as teorias de fonologia de base dinâmica que tomam o gesto articulatório como primitivo de análise tanto de produção como de percepção (Goldstein & Fowler, 2003).

O debate entre Ohala e Fowler ilustra que não há consenso entre os pesquisadores sobre a natureza da percepção da fala. A nossa proposta é a de que esse fato existe porque cada um dos debatedores se situa em um terreno epistemológico que, por sua vez, considera fenômenos distintos como relevantes para sustentar cada uma de suas propostas teóricas, a saber: enquanto Ohala sustenta seus argumentos com base em evidências fonológicas de natureza acústica, Fowler salienta os aspectos articulatórios presentes na percepção e produção da fala.

Palavras-chave: teorias de percepção da fala: audição; gesto pretendido; gesto articulatório; Teoria Motora da Percepção da Fala; Teoria do Realismo Direto da Percepção da Fala; natureza da percepção da fala.

ABSTRACT

The major goal of this thesis is to consider three different ways to postulate the nature of speech perception. On Chapter 1, we present that speech perception can be based on acoustic or articulatory properties. Acoustic-based account claims the existence of acoustic invariant properties that make the categorization of distinctive units of speech signal able. Articulatory-based account takes the articulatory gesture as a perceptual unit. The perceptual unit can be abstract (cf. Motor Theory of Speech Perception – Liberman & Mattingly, 1985) or real (cf. Direct-Realistic Theory of Speech Perception – Fowler, 1986 and following).

The specific goal (Chapter 2) is to briefly narrate the way speech perception theories develop. In this historical course, we link the postulation of perceptual primitives (both of acoustic and articulatory basis) to the new phonological primitives. Namely, we will try to verify if the new perceptual primitives adoption is linked to the epistemological choices proposed by phonological theories.

In order to exemplify our proposal, we present (Chapter 3) a debate occurred in 1996 between Carol Fowler and John Ohala about the speech perception basis. The main interest on this debate is that it is not developed. Once that it counts with one text published by each of the participants, we argue that the “death” of discussion indicates that speech perception theories build their postulates over different epistemological fields. This characteristic makes the closure of the debate impossible.

We are concerned about the speech perception basis debated by Fowler and Ohala because it was not until the Structuralism theory that this question was discussed. To the structuralists, speech perception was the same as audition, in other words, researchers only considered data which was distinctive in a speech corpus. Apparently, the problem emerges when Liberman and colleagues (cf. Liberman, 1957) try to find what make the phonemic categories differentiate from each other. Therefore, they have searched for the responsible acoustic clues which distinguish a [p] sound from a [d] one with the Pattern Playback (an instrument that converts a hand-made spectrogram into sound). However, the results show that there is not a direct connection between the acoustics stimulus and the articulation used to produce that sound. This suggests that the articulation guides perception. In this proposal, the invariant features are articulatory, because the acoustic

signal has differences within phonemic categories due to the vocalic context.

Our research shows that the perceptual primitives follow the primitives proposed by the phonological theories, namely: there is no speech perception theory in the Structuralism. However, influenced by the Jakobson, Fant & Halle's Distinctive Features Theory (1952), the first researches try to find the acoustic perceptual invariants; the Motor Theory of Speech Perception (Liberman & Mattingly, 1985) is developed under the Generative framework. This model proposes that the perceptual units are abstract articulations. Influenced by the Generative Phonology (Chomsky & Halle, 1968), the findings of the McGurk Effect (McDonald & McGurk, 1976) and Fodor's Modularity of Mind (1983), the Motor Theory proposes that the perceptual unit is an intended gesture; lastly, there is the Direct-Realistic Theory of Speech Perception proposal (Fowler, 1986 and following). This theory is completely influenced by the evolutionist framework of Gibson (1966). It converges to the coming frameworks that takes phonology as a dynamic discipline and which assume the articulatory gesture as a primitive that serves both speech production and perception (Goldstein & Fowler, 2003).

The debate between Ohala and Fowler (1996) illustrates that there is no consensus about the speech perception basis among researchers. Our proposal is that lack of consensus among researchers exists because each one of the debaters is allocated in an epistemological field that considers different facts as relevant to support their own theoretical proposal, namely: Ohala supports his arguments on acoustic-based phonological evidences and Fowler underlines the articulatory characteristics that integrate speech production and perception.

These are the issues discussed in this thesis.

Key-words: Speech perception theories; audition; intended gesture; articulatory gesture; Motor Theory of Speech Perception; Direct-Realistic Theory of Speech Perception; speech perception's basis

RESUMEN

El principal objetivo en esta tesis doctoral es presentar tres maneras de tomar los primitivos de la percepción del habla. En el primer capítulo presentaremos que la percepción del habla puede ser considerada como audición o articulación. Si uno la toma como audición, se busca la existencia de rasgos acústicos invariantes que establecerían la categorización de las unidades distintivas en el señal de habla. Ya si uno la toma como articulación, ocurre la adopción del gesto como el primitivo de la percepción. Eso puede ser considerado de base abstracta (como sugiere la Teoría Motora de la Percepción del Habla – Liberman & Mattingly, 1985) o real (Fowler, 1986 e siguientes).

El objetivo específico (disponible en el Capítulo 2) es realizar una reconstrucción histórica acerca de la manera como esas teorías son formuladas. Ahí, relacionamos la proposición de primitivos de la percepción con la emergencia de teorías fonológicas que adoptan nuevos primitivos de análisis. Eso es, intentamos comprobar que la adopción de nuevos primitivos para la percepción está relacionada a las opciones epistemológicas hechas por las teorías fonológicas.

Como ejemplo de nuestra hipótesis en este trabajo, presentamos(en el Capítulo 3) un debate que ocurrió el 1996 entre Carol Fowler y John Ohala acerca de la naturaleza o acústica o articulatoria de la percepción del habla. Estamos interesados en ese debate porque él no desarrolla. Como solo hay un texto de cada uno de los participantes, argumentamos que la “muerte” de la discusión enseña que las teorías de percepción se construyen en áreas epistemológicas distintas. Eso imposibilita que se cierre el debate .

Es importante traer ese debate para nuestros días, porque esa cuestión permanecía olvidada. En el tiempo de la Lingüística Estructural, no se discutía cual era la base de la percepción. Para ellos, la percepción era como la audición, o sea, solamente se consideraban relevantes los datos que los expertos tomaban como distintivos en un corpus. Aparentemente, el problema emerge por causalidad cuando Liberman y sus colaboradores (Liberman, 1957) intentaban ubicar los rasgos distintivos de las categorías fonológicas. Para hacerlo, utilizaban el *Pattern Playback* (un aparato que hacía la conversión de espectrogramas dibujadas a mano en sonidos) para determinar cuales eran los rasgos acústicos que hacían la diferenciación entre [p] y [d], por ejemplo. Sin embargo, los resultados no presentaban una relación directa entre los datos acústicos y

los articulatorios. Para esos investigadores, los datos sugerían que la articulación conducía la percepción, pues no hay variaciones en los datos articulatorios y los hay en la señal acústica.

Nuestra investigación enseña que los primitivos de la percepción siguen los primitivos propuestos en las teorías fonológicas, por ejemplo: en la Lingüística Estructural no había una teoría de percepción. Sin embargo, las investigaciones intentaban ubicar lo que había de invariante en la señal acústica por influencia de la Fonología Estructural de Jakobson, Fant & Halle (1952); en la Generativa, la Teoría Motora de la Percepción del Habla (Liberman & Mattingly, 1985) propone primitivos articulatorios abstractos. Su versión revisada (y más fuerte) es influenciada por la propuesta de la Fonología Generativa de Chomsky & Halle (1968), los hallazgos acerca del *Efecto McGurk* (McDonald & McGurk, 1976) y la teoría modular de la mente de Fodor (1983); finalmente, se propone la Teoría del Realismo Directo de la Percepción del Habla (Fowler, 1986 y siguientes) por influencia de la teoría de percepción de Gibson (1966). En ese enfoque se desarrolla juntamente una teoría dinámica de fonología (la Fonología Articulatoria) que adopta el gesto articulatorio como la base de la producción y percepción del habla (Goldstein & Fowler, 2003).

El debate entre Ohala y Fowler enseña que no hay consenso entre los investigadores acerca de la naturaleza de la percepción del habla. Nuestra propuesta sugiere que la falta de consenso existe porque cada uno de los comentaristas se ubica en una área epistemológica distinta. Eso hace con que ellos consideren fenómenos distintos como relevantes para la sustentación de sus teorías. Por ejemplo, mientras Ohala sostiene su argumentación con evidencias fonológicas de base acústica, Fowler prefiere los rasgos articulatorios que se hallan en la percepción y la producción del habla.

A esas cuestiones esta tesis intenta contestar.

Palabras-clave: teorías de percepción del habla; audición; gesto intentado; gesto articulatorio; Teoría Motora de la Percepción del Habla; Teoría del Realismo- directo de la Percepción del Habla; naturaleza de la percepción del habla.

PREFÁCIO

Há 10 anos ocorreu o meu primeiro contato com os estudos fônicos e, por isso, permito-me utilizar aqui todos os clichês possíveis de um prefácio de tese. Não tenho medo de dizer que essa disciplina foi *amor à primeira vista*. Abduzido de imediato pelos sons, articulações e representações diversas, debrucei-me desde cedo (graças ao convite da professora doutora Adelaide Silva) sobre os estudos da fala.

Embora o tom piegas seja permitido nessas circunstâncias de fim de uma etapa, ele se faz necessário justamente para explicitar a natureza deste trabalho. Posso dizer que este trabalho era uma crônica anunciada, uma vez que ela foi crescendo e tomando corpo de maneira quase que sozinha.

Explico. Este trabalho não era para ser uma pesquisa sobre as teorias de percepção da fala. Minha motivação inicial era tentar descobrir a importância perceptual do elemento vocálico produzido junto ao *tap*. Para minha mente ainda um tanto quanto ingênua, pensei que essa seria a discussão que começaria a responder os pontos que fui deixando de lado nos meus primeiros trabalhos. Essa perspectiva fica clara ao se resgatar o título do meu pré-projeto de ingresso no curso de doutorado da Universidade Federal do Paraná (UFPR) no fim de 2008: *Entre o fonético e o fonológico: o problema da percepção do tap no PB*.

Nesse título, também fica evidente que o *fruto não cai longe do pé*, pois desde o primeiro contato com os aspectos fônicos da linguagem (nas aulas de fonética dos anos da graduação) era evidente uma discussão nada ingênua sobre a delimitação das disciplinas fonética e fonologia. Essas discussões sobre zonas de fronteira, primitivos de análise e limites de um determinado modelo estiveram presentes durante toda a minha formação na graduação e foram acentuadas durante o mestrado. No entanto, faltava-me fôlego para desenvolver tais discussões, embora fossem as que eu gostaria de ter realizado durante aquela etapa.

Para então tentar dar um norte ao meu trabalho de doutorado, durante o primeiro ano me detive a cumprir os créditos e a responder uma única pergunta: *o que é percepção da fala?* Confesso que aí se iniciou uma viagem que culminou neste texto.

Ao me deparar com poucas teorias sobre a percepção da fala, fui obrigado a tentar organizar tais propostas dentro dos arcabouços teóricos da linguística. Aí se iniciou meu namoro com a história e a filosofia da linguística. A essa altura eu já tinha mudado o título

do meu trabalho mais uma vez: *Sobre a natureza da percepção da fala e a demarcação das teorias*.

Aqui fica evidente (pelo menos para mim) que eu estava andando sobre novos terrenos e só consegui enxergar isso por uma pequena pergunta que a professora doutora Lígia Negri me fez: *mas o seu trabalho vai ser sobre percepção ou sobre as teorias de percepção?* Pergunta singela? Não! Definitivamente, não! Aceitar que a tese falava sobre as teorias era assumir que as respostas tão esperadas sobre a percepção do *tap* teriam que ser postergadas. Enfrentei o problema! Aqui está a tese e mais uma vez com um título novo (e espero que definitivo): *sobre teorias de percepção da fala*.

Três coisas podem ser ditas sobre esse título. Primeiro, consegui superar a minha própria tendência de seguir o gênero acadêmico de nomeação de trabalhos que geralmente faz uso de títulos longos, incompreensíveis, sem contar (é claro) com o uso devoto dos dois pontos. Segundo, esse título salienta a natureza desta descrição e reflexão que é antes de tudo teórica e histórica. Terceiro, não se trata (é claro) de uma revisão exaustiva de todas as teorias de percepção da fala. Trata-se de um recorte sobre as que julguei de maior impacto dentro de cada “período” da história das teorias linguísticas.

Em suma, o que eu espero ter deixado claro é que a intenção inicial não era a de fazer um trabalho de natureza histórica. Talvez o tom mais didático dos dois primeiros capítulos revele o meu próprio processo de aprender sobre a percepção da fala. E a nebulosidade do terceiro capítulo pode refratar minhas dificuldades frente a questão levantada.

Um último esclarecimento: a epígrafe colocada aqui espelha algumas reflexões realizadas durante todo o processo de escritura da tese. Primeiro, o título da música mostra minha busca por calma e tranquilidade para conseguir responder as perguntas a que me fiz. Controlar a ansiedade é uma tarefa constante. Segundo, ela ilustra também uma mudança de objeto de estudo. O debruçar sobre a história da linguística tem me proporcionado um distanciamento teórico interessante e, conseqüentemente, me fez amadurecer com relação às minhas preferências e militâncias teóricas (que – para quem me conhece – sempre foram veementes e apaixonadas). Por fim, Paul Simon tem me acompanhado durante todo esse processo. Seu disco de 1972 é a trilha sonora deste trabalho. Qualquer semelhança com Douglas Adams e seu *Guia* não é mera coincidência.

INTRODUÇÃO

Sem nenhuma exceção¹, os seres humanos nascem sem falar. Em contrapartida, desde muito cedo as crianças já ouvem os sons, sentem os cheiros e as texturas das coisas e veem objetos e pessoas do seu ambiente. Essa capacidade de perceber seu ambiente e os seres e as coisas presentes tem sido considerada fundamental para o desenvolvimento linguístico das crianças independentemente do viés teórico de aquisição de linguagem adotado pelos pesquisadores da área².

Considerando essa gama de sensações em que os seres humanos estão imersos, é sobre a audição que os estudos sobre a percepção da fala vão se debruçar. Não se trata de uma questão nova, pois a pesquisa em linguística já dedicou algum tempo (por volta de 50 anos se considerarmos os primeiros trabalhos de Liberman e colegas nos Laboratórios *Haskins*³ como sendo o pontapé inicial) sobre como e o que percebemos nos sons da fala.

A título de exemplo sobre a natureza da percepção da fala, podemos dizer que até o Estruturalismo não se questionava qual era a sua natureza. Para os estruturalistas, a percepção era o equivalente à audição, i.e., considerava apenas os dados percebidos como distintivos ao se analisar um *corpus*. No entanto, o problema surge por acaso quando Liberman e colegas (cf. Liberman, 1957⁴) tentavam encontrar o que havia de distintivo entre as categorias fonêmicas. Para isso, eles buscavam (com a ajuda de um aparelho – o *Pattern Playback* – que convertia

¹ Até mesmo nos casos mais extremos da ficção isso não parece acontecer. No conto “O curioso caso de Benjamin Button” (publicado em 1922 no livro “Tales of the Jazz Age”), Scott Fitzgerald narra a bizarra história de um rapaz que nasce velho e vai ficando mais jovem à medida que o tempo passa. No entanto, o protagonista não sabe falar ao nascer.

² No seu célebre “*Da fala à linguagem: tocando de ouvido*”, Albano (1990) mostra que a criança entra na linguagem pelo som. Ela narra ali casos em que as crianças reproduzem expressões inteiras (tal como a vinheta da *Rádio Globo*) mesmo antes de se utilizarem da fala para se comunicar.

³ Os Laboratórios *Haskins* são um instituto de pesquisa (privado e sem fins lucrativos) sobre fala, linguagem, leitura e suas bases biológicas. Foi fundado em 1935 e possui sede, desde 1970, em New Haven (Connecticut – EUA). Também apresenta filiação formal com a Universidade de Connecticut e a Universidade de Yale. Para maiores informações sobre o instituto visite www.haskins.yale.edu.

⁴ As referências de Liberman (1957, 1967) e Liberman & Mattingly (1985) se encontram reunidas no volume Liberman (1996). Decidi manter as datas originais no texto para salientar o percurso histórico do seu trabalho. Contudo, as citações presentes aqui seguem a paginação da coletânea de 1996.

espectrogramas desenhados à mão em som⁵) as pistas acústicas que diferenciavam um [p] de um [d], por exemplo. Contudo, os resultados não apresentavam uma relação biunívoca entre o sinal acústico e a articulação realizada. Para os autores, isso sugeria que a percepção da fala somente era possível se o ouvinte estivesse sendo guiado pela articulação e não pela audição, devido à falta de invariância do sinal acústico⁶.

É a partir desse ponto que o problema se instaura e, com isso, teorias sobre a percepção da fala surgem para tentar explicar os fatos encontrados por Liberman e colegas. Entretanto, somente em 1996 ocorre um “debate”⁷ sobre a natureza da percepção da fala. Embora tenha sido muito breve, a troca de textos entre John Ohala e Carol Fowler será o ponto de chegada para a nossa discussão sobre a natureza da percepção da fala, pois o **objetivo geral** deste trabalho (presente no Capítulo 1) é realizar uma reflexão a partir da apresentação de um panorama sobre as teorias de percepção da fala com relação aos primitivos de análise adotados. A novidade deste trabalho está justamente no fato de que iniciaremos com as abordagens auditivas para depois passarmos para as teorias que adotam primitivos de base articulatória. Isso se faz necessário devido: i) à falta de trabalhos em língua portuguesa que discutam sobre a natureza dos primitivos de percepção da fala e as teorias disponíveis para guiar os experimentos e explicar os dados; e ii) ao crescente interesse pela percepção da fala nos estudos fônicos nacionais⁸.

Como **objetivo específico** (Capítulo 2), estabeleceremos uma ligação entre a ontologia dos primitivos adotados pelas teorias de percepção da fala com o corte epistemológico proposto pelas teorias fonológicas. Realizaremos uma descrição de como as teorias fonológicas propõem novos primitivos de análise que, por sua vez, influenciam a proposição de modelos de percepção. Partiremos dos primeiros achados dos anos 50 nos Laboratórios *Haskins* com os trabalhos de Liberman e

⁵ Para maiores informações sobre o funcionamento do *Pattern Playback* ver a Introdução e o Capítulo 2 de Liberman (1996).

⁶ A semente da busca pela invariância articulatória presente nos trabalhos de Browman & Goldstein (1986 e seguintes), por exemplo, só foi possível devido aos achados iniciais com o auxílio do *Pattern Playback*.

⁷ Ou melhor, a polêmica se torna pública (cf. Dascal, 2006)

⁸ Para um abordagem preliminar sobre a percepção do *tap* em grupos tautossilábicos do PB, consultar Nishida (2011). Sobre a percepção de consoantes finais por falantes brasileiros aprendizes de inglês como L2, ver Albuquerque (2012). E sobre a produção e percepção de aspectos fônicos em crianças com queixa fonoaudiológica, buscar Freitas (2012).

colegas, paralelamente às propostas do Estruturalismo. Passaremos pela formulação da Teoria Motora da Percepção da Fala (Liberman & Mattingly, 1985) considerando seus achados (e os que os influenciaram – e.g., o *Efeito McGurk* apresentado por McDonald & McGurk, 1976) e suas bases teóricas com relação à Fonologia Gerativa Padrão (Chomsky & Halle, 1968) e à modularidade da mente (Fodor, 1983). Por fim, trataremos da Teoria do Realismo Direto da Percepção da Fala (Fowler, 1986) e das propostas mais recentes com relação às abordagens dinâmicas da produção da fala (Browman & Goldstein, 1986 e seguintes).

A hipótese defendida neste trabalho (a ser tratada mais detalhadamente no Capítulo 3) é a de que a escolha do primitivo da percepção é dada epistemologicamente, i.e., a depender da teoria fonológica escolhida é possível contemplar novos fenômenos que, por sua vez, abrem campo para a verificação experimental (a partir de um Programa de Investigação Científica) com relação à produção e percepção da fala. Em outras palavras, será possível verificar que a percepção é tratada auditivamente no Estruturalismo devido às concepções metodológicas (e epistemológicas) propostas para o tratamento do nível fonológico. Já no Gerativismo, há o olhar estritamente mental e simbólico para o tratamento da percepção. Por fim, a abordagem realista e direta para a percepção é proposta em sintonia com os achados experimentais que “implodem” a dicotomia das disciplinas Fonética e Fonologia e com uma teoria de percepção de base ecológica. Em outras palavras, definir o que é percepção da fala depende do que se toma como sendo o nível fônico da linguagem, pois ele pode ser estritamente o domínio acústico (tal como propunha o Estruturalismo), mental (no Gerativismo) ou real (nas abordagens dinâmicas de viés ecológico).

Cabe ainda mencionar que além de traçar historicamente a trajetória de cada uma das teorias de percepção da fala (no Capítulo 2), realizaremos uma discussão nas páginas finais deste trabalho (no Capítulo 3) acerca de um “debate” ocorrido entre Carol Fowler (1996) e John Ohala (1996) para ilustrar (e defender nossa hipótese) que a adoção de um primitivo para a percepção está vinculada à proposição de novos primitivos de análise fonológica. Em outras palavras, se os primitivos perceptuais são determinados pela demarcação epistemológica, é possível comparar as teorias de percepção? Será que cada teoria está “falando” de

um lugar diferente? Nossa proposta é de que cada teoria trata a percepção da fala de maneira distinta, pois suas bases teóricas estão assentadas sobre terrenos epistemologicamente distintos. Seria por este motivo que os debates (tal como o de Fowler e Ohala) não se desenvolvem⁹.

Por fim, esclarecemos que a natureza deste trabalho é puramente descritiva. Ou melhor, estamos tentando reconstruir como se dá a formulação das teorias de percepção sem uma preferência por alguma delas. O caráter linear e cronológico da apresentação pode sugerir uma preferência pelas abordagens mais recentes de base dinâmica. Contudo, isso é apenas um efeito (indesejado) da construção da narrativa¹⁰.

⁹ É importante levar em conta que, por mais que o debate não se desenvolva entre os participantes, podemos considerar que a polêmica permanece, pois não há um “vencedor”. Isso faz com que pensemos que, em realidade, a questão continua e ela pode ser retomada a qualquer momento por qualquer “interessado” no assunto. Esse ponto é importante ter em vista, uma vez que sugeriremos nas considerações finais deste trabalho que os estudos sobre a filosofia e história da linguística (e por que não da ciência?) se beneficiariam com a adoção de um princípio discursivo como o *dialogismo*. Levá-lo em consideração promoveria uma explosão de textos (enunciados) que dialogam com questões colocadas em outros debates e polêmicas. Este não é o nosso ponto principal do trabalho, mas trata-se de um pano de fundo que certamente teremos que desenvolver em trabalhos futuros.

¹⁰ O fato de realizarmos uma descrição não sugere que não temos uma “preferência” teórica. O que queremos deixar claro é que a nossa preferência teórica não está sendo defendida neste trabalho, uma vez que estamos tentando refletir sobre a constituição histórica e epistemológica das teorias de percepção da fala. Contudo, isso não significa que em algum momento não possamos ter deixado “escapar” alguma expressão que produza algum efeito de sentido “a favor” de alguma abordagem. Se as encontrarem, por favor, desconsideram-nas.

CAPÍTULO 1: A NATUREZA DA PERCEPÇÃO DA FALA¹¹

O'Callaghan & Nudds (2009) propõem que a filosofia da percepção deve começar a olhar para dados de percepção sonora. Segundo os autores, historicamente, apenas se considerava a percepção visual e depois se transpunha tais noções aos outros sentidos. Em seu texto, há dois pontos principais: (i) um sobre qual seria a ontologia dos sons; e (ii) outro preocupado com questões sobre o conteúdo da experiência auditiva. Nesse segundo ponto, são realizadas perguntas sobre *“como os sons são experienciados?”*, *“o que e quais propriedades são experienciadas na percepção sonora?”* e *“o que ouvimos quando ouvimos a fala?”*.

Por hora, o nosso maior interesse está nas perguntas sobre a ontologia dos sons. Os autores traçam um breve histórico sobre as diferentes ontologias propostas na literatura filosófica. A sua argumentação parte do ponto de que a natureza do som está ligada à sua experiência (O'Callaghan & Nudds, 2009: p.5). Por exemplo, Locke (cf. O'Callaghan & Nudds, 2009) agrupou os sons, as cores, os sabores e os cheiros como disposições cujas caracterizações os atavam essencialmente à experiência dos sujeitos. Já no século 20 (cf. MacLachlan, 1989 *apud* O'Callaghan & Nudds, 2009: p.5), os sons são subjetivos e privados, i.e., assume-se uma visão representacional sobre os sons uma vez que eles medeiam o acesso auditivo perceptual ao mundo. Em oposição a essa abordagem, há a possibilidade de considerá-la de uma maneira não representacional, i.e., perceberíamos os sons imediatamente, diretamente.

Há ainda a possibilidade de se tomar os sons como propriedades atribuídas às coisas comumente tomadas como a sua fonte sonora. Por exemplo, os sons de um sino facilmente são reconhecidos pelas pessoas e, ao ouvir seu som, elas o atribuem à sua fonte sem maiores problemas. Pode-se também considerar os sons como indivíduos. Desta forma, eles carregam traços sensíveis (i.e., que podem ser ouvidos, sentidos, percebidos) como *pitch*, timbre e intensidade particulares.

¹¹ Neste primeiro momento, apenas apresentaremos as diferentes maneiras de se considerar a natureza da percepção da fala.

Outro olhar sobre os sons reside no fato de eles serem persistentes, de terem duração no tempo. Isso faz com que eles sejam tidos como objetos. Seguindo esta noção também é possível entender o som como eventos. Desta maneira, não seriam apenas vibrações, seriam vibrações de algum objeto que provoca a perturbação de um meio. Tais vibrações (i) não são idênticas às vibrações do objeto, pois (ii) resultam da interação de múltiplos objetos e (iii) precisam de um meio para se propagar (esse evento não ocorre no vácuo, é claro). Por fim, pode-se considerar o som como ondas de pressão longitudinal que passam através de um meio elástico como o ar, água ou metal. As ondas propagam de sua fonte até seus observadores, i.e., “os sons seriam apenas ondas viajantes” (O’Callaghan & Nudds, 2009: p.6).

Diante dessa vasta gama de possibilidades frente à natureza dos sons, os autores acabam dividindo tais abordagens em dois tipos: (i) os tratamentos baseados em ondas [*wave based account*]; e (ii) os baseados nas fontes [*source based account*]. Essa divisão implica em 3 tipos de conteúdos da experiência auditiva: (i) ouvimos imediatamente os sons e seus atributos (*pitch*, timbre e intensidade); (ii) ouvimos sons e sua fonte; e (iii) ouvimos todas as coisas que estão no ambiente além do som e sua fonte.

Assim, é possível notar que não é tão simples estipular uma ontologia para a percepção da fala. Parece haver tantas ontologias quanto experiências possíveis. Com isso, tomo a divisão proposta pelos autores (tratamentos baseados na fonte e em ondas) para traçar algumas diferenças entre as teorias de percepção da fala com relação à natureza do primitivo de análise.

Neste capítulo, apresentaremos duas correntes: uma de viés auditivo, baseada em ondas (seções 1.1, 1.2 e 1.3); e outra articulatória, baseada na fonte. Como vocês perceberão, dentro da corrente articulatória, há duas teorias formuladas (a *Teoria Motora da Percepção da Fala* – Liberman & Mattingly, 1985, seção 2.1.1 – e a *Teoria do Realismo Direto da Percepção da Fala* – Fowler, 1996, seção 2.2.1). O fio condutor deste capítulo reside nas diferentes ontologias adotadas pelas teorias da percepção da fala.

1 COMO AUDIÇÃO

Quem de nós nunca presenciou ou ficou sabendo de uma anedota que envolva alguma “trapalhada” linguística? Com certeza, muitos de nós as presenciou e não é preciso ser um linguista para notá-las. Quantas vezes nos deparamos com pessoas de outras regiões do país (ou, simplesmente, praticantes de outros dialetos de uma mesma localidade) que dizem sem medo algum “Nossa como *você fala engraçado!*” ou “*Como você fala diferente!*”. Tais casos são mais frequentes do que imaginamos e a maioria desses casos pode ser um argumento em direção ao fato de que os seres humanos percebem diferenças acústicas na fala.

Voltando às “trapalhadas” linguísticas, podemos dizer que o anedotário é, na sua grande maioria, composto por generalizações realizadas por crianças ou por estrangeiros tentando falar outra língua.

Por exemplo, presenciei mais de uma vez brincadeiras feitas por um amigo a uns intercambistas japoneses que estudavam na Universidade Federal do Paraná (UFPR), lá por meados de 2002. Esse amigo estava ensinando os alunos japoneses a falar *tchutchuca* (gíria para se referir à mulher bonita) e não havia o que os fizesse falar com a acentuação desejada, i.e., ao invés da versão paroxítona esperada (tchuTCHUca), os japoneses produziam a palavra como uma oxítona (tchutchuCÁ). Nessa ocasião, a maioria das pessoas envolvidas na “brincadeira” comentava que eles não ouviam a diferença entre as versões paroxítona e oxítona e acabavam reproduzindo a palavra no padrão acentual japonês, ou seja, oxítono. Em contrapartida, nós, falantes de português, notamos a diferença acentual, pois a posição do acento é distintiva na nossa língua. A título de exemplo, temos a boa e velha tríade mattosiana *sábia*, *sabía* e *sabiá* (Câmara Jr, 2001 [1970]). No exato momento em que ocorria esse fato, não nos dávamos conta de que em realidade os japoneses podiam sim estar fazendo alguma distinção entre oxítonas e paroxítonas, só que essa diferença talvez não fosse suficiente para gerar uma distinção auditiva saliente aos nossos ouvidos guiados pela fonologia do português.

Esses fatos podem embasar propostas de que percebemos apenas as distinções fonologicamente relevantes, i.e., as distinções que produzem contrastes na nossa língua.

No entanto, a questão não para aí. Aqui vai um exemplo com relação às segmentações que podem ser feitas no sinal acústico. Já faz algum tempo que eu ouvi a estória de um menino que pergunta à mãe o que é *blado*¹². A mãe, é claro, pergunta o que ele quer dizer com isso. Ele responde que quer saber onde é que fica o *blado*, pois alguém havia dito perto dele que “o céu está no *blado*” [nublado]. Como a criança conhece muitas sequências possíveis da sua língua e *no blado* poderia ser algo parecido a *no parque*, *no supermercado* ou *no shopping*, ela segmentou a sequência sonora desconhecida de *nublado* como uma forma conhecida.

Temos, neste caso, mais uma vez, um argumento de que as pessoas segmentam o sinal acústico a partir de generalizações com relação à sua própria língua, seguindo características fonotáticas¹³, por exemplo.

É diante desses tipos de dados que temos uma primeira abordagem com relação à percepção da fala. Trata-se de uma abordagem auditiva sobre o sinal acústico. Apresentaremos, nas seções seguintes (1.1, 1.2 e 1.3), três linhas de pesquisa que consideram as propriedades acústicas da fala como pistas para a sua percepção.

1.1 Dois estágios nos processos perceptuais

No início dos anos 70, uma das abordagens disponíveis sobre a percepção da fala tentava estabelecer experimentalmente uma relação dela com a memória. Pisoni (1973, 1974), Crowder (1973), Carney (1977) e Oden & Massaro (1978) elaboraram experimentos justamente para testar essa hipótese. Nas páginas seguintes, apresentaremos os trabalhos de Pisoni (1973, 1974).

Nesse momento, veiculava-se (principalmente nos estudos dos Laboratórios *Haskins*) a noção de que haveria mecanismos especiais para a percepção da fala, i.e., a percepção da fala seria diferente da percepção de outros sons (tópico a ser

¹² Créditos à professora Maria Laura Mayrink (via professora Adelaide Silva) sobre essa história contada em 1990 na disciplina “Práticas de leitura e produção de textos”.

¹³ Exemplos do mesmo tipo (segmentações a partir de uma “neutralização” fonética como harmonia vocálica) são utilizados aos montes na literatura e na música. Basta olhar para a faixa de abertura do álbum *Muito Pouco* de Paulinho Moska: *Devagar, divagar ou de vagar*.

abordado na seção 2). Essa proposta era sustentada pelos argumentos de que as consoantes são percebidas categoricamente e de maneira não linear. Por exemplo, quando os estímulos são apresentados em pares em um dado experimento, os ouvintes conseguem perceber essas duplas como distintas, mas como pertencentes a uma mesma categoria fonêmica. Isso ocorre justamente pelo fato de as consoantes não serem percebidas de maneira linear, pois a diferenciação fonêmica de uma consoante a outra é dada a partir de um dado ponto do *continuum* de F2, por exemplo. Em contrapartida, as vogais são percebidas continuamente como ocorre com outros sons. Os ouvintes conseguem perceber muito mais categorias intrafonêmicas nas vogais do que nas consoantes¹⁴.

Diante das diferenças na maneira como são percebidos os sons da fala, propõe-se que haveria uma Percepção Categórica (ou Fonética) e uma Percepção Contínua (ou Auditiva).

Essa proposta foi feita por Fujisaki & Kawashima (1969, 1970 *apud* Pisoni, 1973: p.254). Pisoni comenta que

“eles [Fujisaki & Kawashima] sugerem que a distinção categórica-contínua na percepção da fala pode ser relacionada ao grau que separa os componentes da memória auditiva e fonética empregados durante o processo de decisão da tarefa de discriminação”¹⁵.

Eles explicam que, quando dois estímulos de categorias diferentes são apresentados, a sua discriminação é realizada tendo por base uma *memória de curto prazo*. Os sons, então, são identificados como pertencentes a categorias diferentes ou iguais por uma decisão binária.

Essa diferença na maneira como as unidades são percebidas mostra que a percepção categórica tem que ser contrastada com dados de percepção contínua, uma vez que esses dados mostram o funcionamento de uma memória de curto prazo. Desta forma, os ouvintes têm que utilizá-la para comparar as propriedades auditivas armazenadas e, assim, eles podem fazer uma identificação comparada (e não absoluta) baseada nas propriedades acústicas dos estímulos.

¹⁴ A título de exemplo é preciso lembrar o professor Higgins conversando com o coronel Pickering sobre a quantidade de vogais produzidas numa gravação que eles ouvem ainda no início de *My Fair Lady*.

¹⁵ “They suggest that the categorical-continuous distinction in speech perception may be related to the degree to which separate auditory and phonetic memory components are employed in the decision process during discrimination” (Pisoni, 1973: p.254).

Diante desses fatos, o estudo de Pisoni (1973) tem por objetivo verificar se consoantes e vogais são percebidas de maneira distinta com relação ao grau de discriminação do código distinto da memória auditiva e fonética. Será que a percepção categórica usa de maneira diferente a memória de curto prazo? A ideia do experimento é apresentar dois estímulos com intervalos distintos entre eles e os informantes terão que dizer se os estímulos são iguais ou diferentes. Assim, será possível verificar como os ouvintes utilizam a memória com relação a estímulos que variam entre categorias e dentro delas.

No experimento, foram estipuladas três variáveis independentes: 1) condições do estímulo (4 níveis: 2 classes de vogais e 2 de consoantes); 2) comparações de estímulos (2 níveis: dentro de categorias fonéticas e entre categorias fonéticas); e 3) atraso de intervalo (0,0; 0,25; 0,50; 1,0; e 2,0s). 16 informantes participaram do experimento.

Os estímulos da oclusiva sonora /bœ/~/dœ/ foram sintetizados seguindo os seguintes parâmetros: os valores de F1, F2 e F3 da vogal eram 743Hz, 1620Hz e 2862Hz, respectivamente; a consoante apresentava vozeamento de 40ms em 150Hz. O experimento partia de /b/ (com transição de 40ms; F2 de 1232Hz e F3 de 1695Hz) e ia (após sete variações iguais) até /d/ (com F2 de 2180Hz e F3 de 3195Hz). O F0 de todos os estímulos era de 120Hz.

Já os estímulos que variavam com relação ao vozeamento (/ba/ e /pa/) foram sintetizados como sílabas de 300ms com a vogal /a/ no núcleo silábico (F1, F2 e F3 de 769Hz, 1232Hz e 2525Hz, respectivamente). A vogal tinha 250ms de duração. O que variou nestes estímulos foi o VOT, que partia de 0 até 60ms. A diferença de cada estímulo era de 10ms.

A parte do experimento com relação às vogais apresentou dados de vogais com estado estacionário longo e curto. Nos dois conjuntos de dados, foram sintetizados os 3 primeiros formantes de um /i/ até um /I/. No contexto longo, as vogais tinham 300ms de duração e, no curto, 50ms. As vogais curtas foram inseridas no experimento pois Fujisaki e Kawashima (1969, 1970) reportaram que vogais breves tendem a ser percebidas categoricamente¹⁶.

¹⁶ Pisoni não comenta o motivo pelo qual Fujisaki & Kawashima propõe que as vogais curtas tendem a ser percebidas categoricamente. Uma possível explicação seria a de que nas vogais longas há

Para cada uma dessas condições foram rodados dois experimentos: (1) teste de identificação absoluta; e (2) teste de reconhecimento de memória por comparação atrasada. Em (1), os informantes tinham que decidir entre as categorias previamente estipuladas pelos pesquisadores e, em (2), os informantes tinham que dizer se eram iguais ou diferentes.

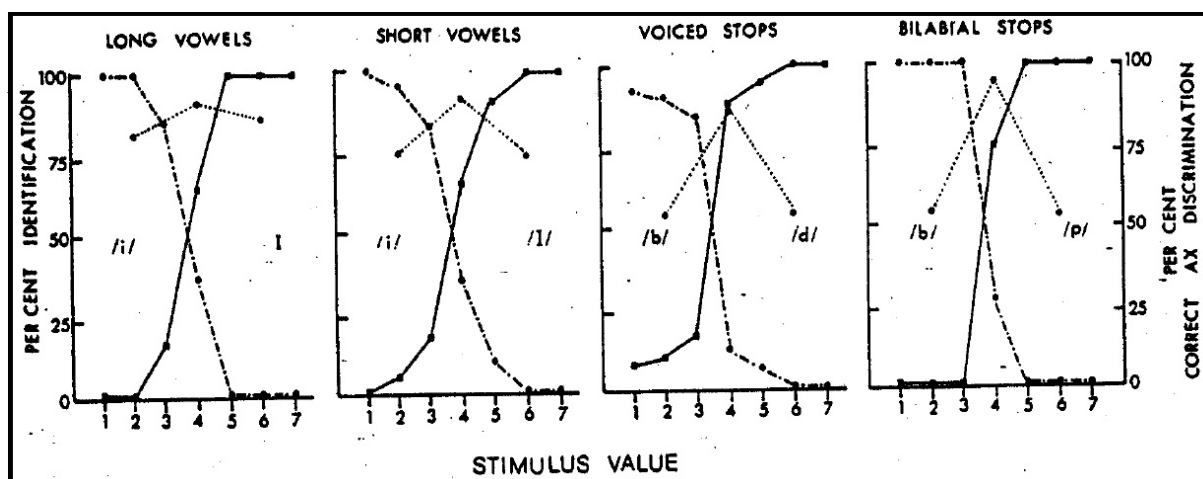


Ilustração 1: Resultados do teste de identificação absoluta (cf. Pisoni, 1973: p.256).

No experimento de identificação absoluta, os sujeitos dividiram o contínuo sonoro em duas categorias (dois segmentos fonéticos distintos). A performance de discriminação é melhor entre categorias fonéticas do que dentro das categorias fonéticas, i.e., à medida que as diferenças entre os estímulos aumentam, aumenta a sua identificação como outra categoria. No entanto, a discriminação dentro da categoria é perto de chance para consoantes e acima dela para as vogais. Também há melhores resultados para as vogais longas do que para as breves.

Com relação ao teste de reconhecimento de memória por comparação, temos dois procedimentos avaliadores: (1) estímulos discriminados como diferentes (representados pela fórmula $P("D"|D)$); e (2) estímulos iguais e diferentes para evitar vieses (representados por d' , que indica a força da memória nas tarefas de reconhecimento).

mais tempo para uma variação acústica e articulatória e, por isso, tenderiam a ser percebidas de outra forma.

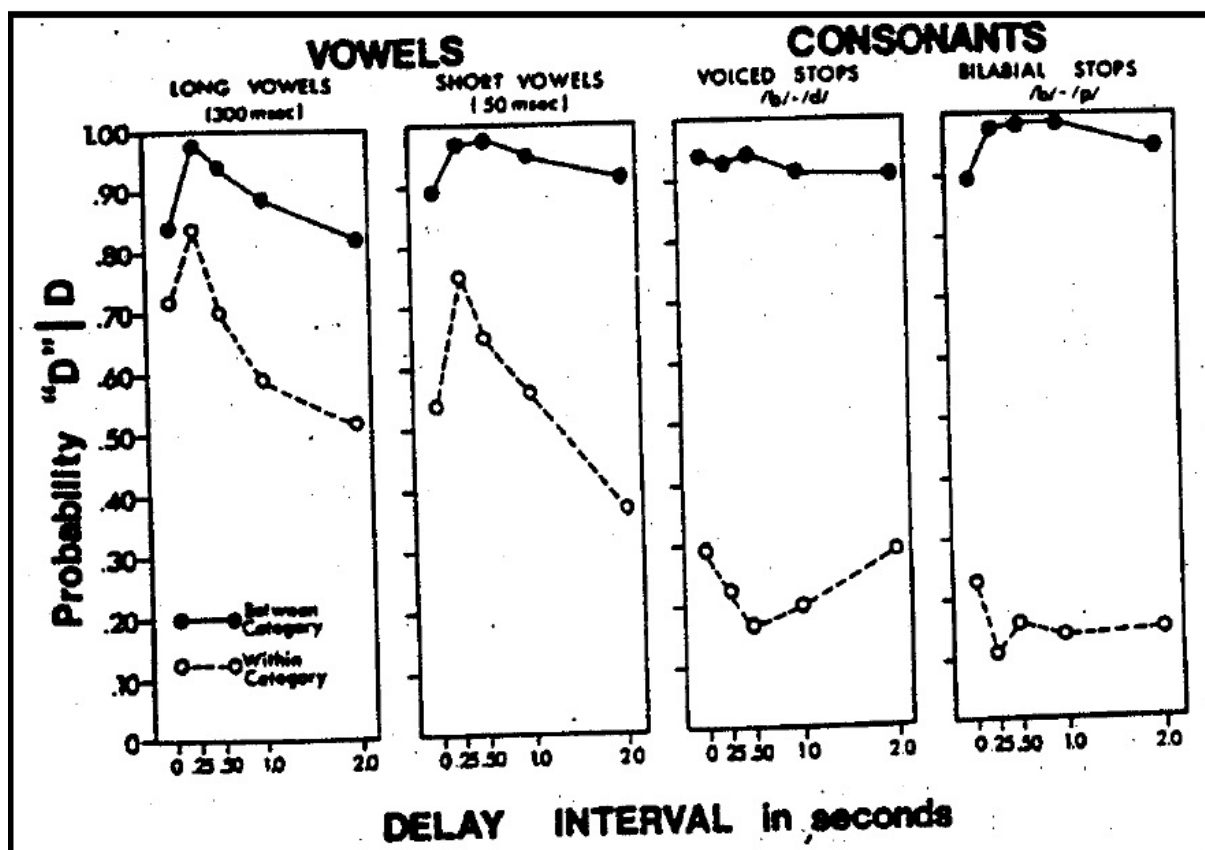


Ilustração 2: Resultados para o procedimento (1) (cf. Pisoni, 1973: p. 256).

Na ilustração 2, notamos que, entre categorias, há estabilidade nas discriminações quando se tem intervalo. Embora isso não ocorra para vogais longas, os testes estatísticos mostraram que isso não apresenta significância. Já dentro de categorias, há diferenças marcadas nas discriminações de consoantes e vogais. As consoantes apresentam valores baixos e não há relação à duração do intervalo. Para as vogais, ao contrário, os valores são significativamente altos e relacionados a mudanças nos intervalos. Por exemplo, o resultado para o intervalo de 0.25s é diferente dos outros. É nele que há maior discriminação.

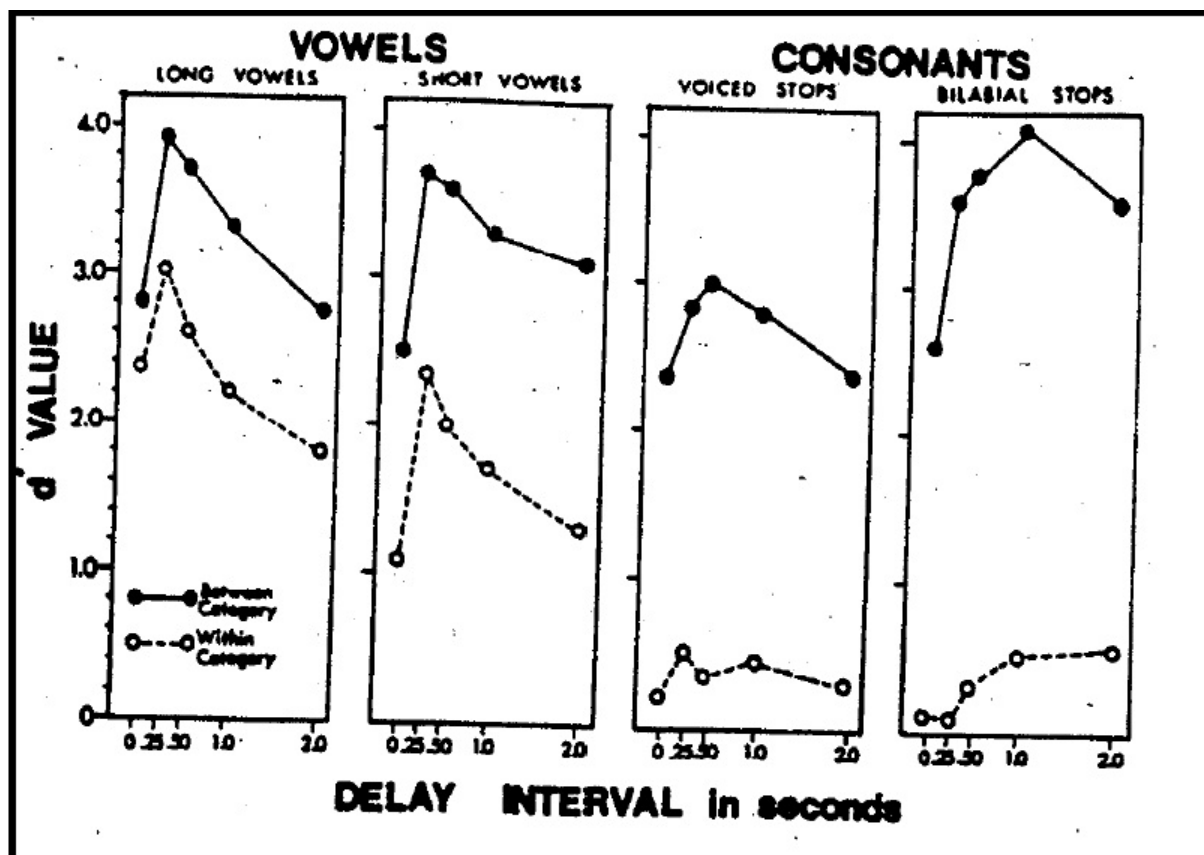


Ilustração 3: Resultados para o procedimento (2) (cf. Pisoni, 1973: p.257).

Considerando o segundo procedimento de avaliação (Ilustração 3), notamos que a discriminação cai nos dois ambientes de avaliação a partir de 0.25s. No que tange às diferenças entre consoantes e vogais, percebemos que a discriminação de vogais é mais afetada pelo aumento do intervalo e as maiores diferenças também são obtidas dentro das categorias.

Em suma, os resultados mostram uma diferença no uso da memória auditiva de curto prazo para a identificação de consoantes e vogais. Com isso, eles sugerem que há dois componentes de memória, devido aos resultados de preponderância (P("D"|D)). Dentro de categorias fonéticas, há a utilização da memória auditiva de curto prazo, uma vez que ela facilita a discriminação de vogais dentro de categorias e contribui pouco para as consoantes. Já entre categorias fonéticas, a memória fonética de curto prazo não interfere, pois a identificação é alta tanto para vogais quanto para consoantes e não é afetada pelo aumento do intervalo de tempo.

Com relação aos resultados para a força da memória de curto prazo (d'), o

valor de d' diminui com o aumento do intervalo de tempo. A falta de uma condição de estímulo significativa pelo intervalo de tempo com o valor de d' sugere que a discriminação de consoantes também é influenciada em alguma medida pelo intervalo de tempo.

Diante disso, esses resultados mostram que a diferença de uso na memória de curto prazo para vogais e consoantes parece ser uma simplificação. A informação acústica necessária para a discriminação de duas consoantes fisicamente diferentes, mas foneticamente idênticas, não está disponível para o uso na discriminação, mesmo em intervalos curtos. Por que essa informação não está disponível para a discriminação?

Fujisaki & Kawashima (1969, 1970) explicam que as pistas para a identificação de consoantes são muito breves para serem estocadas na memória de curto prazo. Desta forma, a categorização de vogais não é absoluta e essas propriedades auditivas estão presentes para serem utilizadas nas discriminações (devido aos resultados acima de chance para as discriminações dentro de categorias). Já a categorização de consoantes é absoluta e talvez seja preciso um codificador especial (sugestão de Liberman, Cooper, Shankweiler & Studdert-Kennedy – 1967), devido aos resultados próximos à chance. Com esses dados não se sabe qual seria a natureza desse codificador (de base acústica ou articulatória) ou se se trata de uma limitação do sistema auditivo. Por fim, há diferenças no uso de pistas de vogais e consoantes pela memória auditiva de curto prazo de modo que ela é melhor para as pistas acústicas de vogais do que de consoantes.

Em 1974, Pisoni tenta desenvolver melhor a ideia de que haveria estágios no processamento da fala. Ele parte da noção de que há percepção categórica, uma vez que um sujeito pode identificar a categoria fonética de um som da fala, mas não consegue discriminar entre sons da fala (selecionados como acusticamente diferentes) como de uma mesma categoria fonética. (cf. Liberman e colegas, 1967; Liberman, 1970; Pisoni, 1971). Ao que parece, o fenômeno é exclusivo a algumas classes de sons. A explicação oferecida neste momento é que, devido à complexidade do sinal da fala e aos experimentos (informantes tendo que solucionar questões em poucos segundos), os sujeitos são forçados a utilizar uma representação fonética ou um código fonético ao invés de utilizar a memória auditiva

de curto prazo para desempenhar as tarefas.

O autor comenta¹⁷ que

“recentemente, Pisoni (1973) sugeriu que a informação auditiva dos estágios iniciais do processamento perceptual dos sons da fala tendem a ser perdidos mais rapidamente do que a informação de estágios mais tardios da análise. Portanto, a informação acústica subjacente a certas classes de sons da fala, tais como consoantes oclusivas, parecem estar indisponíveis para o uso nas tarefas de discriminação subsequente” (Pisoni, 1974: p. 285)¹⁸.

No entanto, há a possibilidade de as informações iniciais estarem disponíveis pelo menos por um breve período de tempo. Com isso, surge a hipótese de que há estágios no processamento dos sons da fala. Pisoni propõe o seguinte modelo de processamento da fala.

¹⁷ Todas as traduções que forem minhas serão acompanhadas dos seus respectivos originais em nota de rodapé.

¹⁸ “Recently, Pisoni (1973a) has suggested that auditory information from the earliest stages of perceptual processing for speech sounds tends to be lost more rapidly than information underlying certain classes of speech sounds such as stop consonants appears to be unavailable for use in subsequent discrimination” (Pisoni, 1974: p.285).

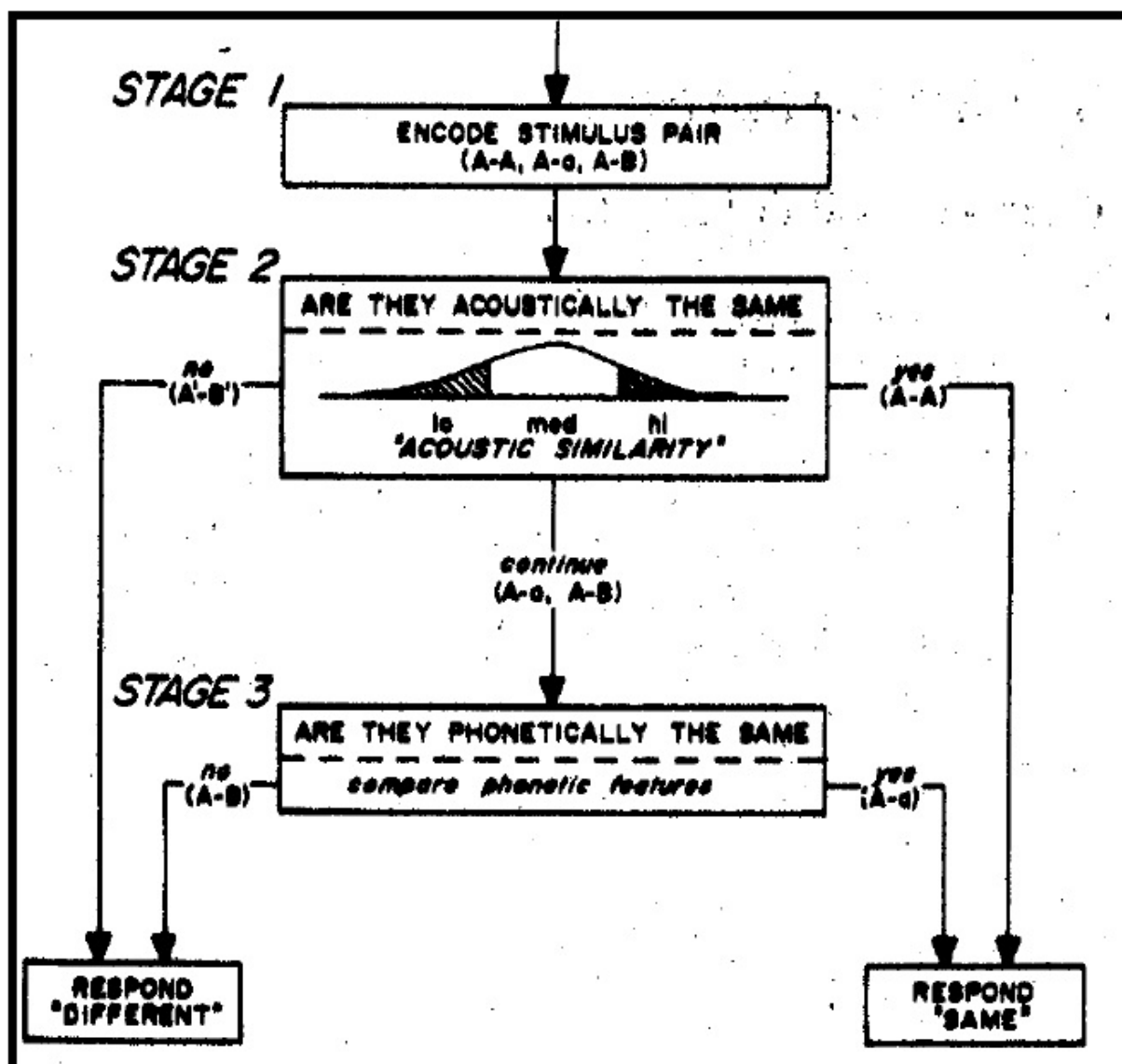


Ilustração 4: Proposta de Pisoni (1974: p.286) para os estágios de processamento da fala.

Assumindo esse modelo de processamento da fala, o autor decide trabalhar com a relação entre as informações auditivas e fonéticas envolvidas na percepção da fala. Especificamente, ele quer saber se os ouvintes conseguem responder a diferenças acústicas entre sons da fala percebidos categoricamente ou se eles conseguem processar esses sons apenas tendo uma abstração fonética como base. O trabalho se baseou em medidas de tempo de reação em um experimento em que os informantes terão que dizer se os estímulos são iguais ou diferentes. Os estímulos utilizados serão do tipo AA (i.e., iguais), Aa (sendo que *a* é uma variação de A) e AB (sendo B de outra categoria).

O experimento contou com a participação de 9 informantes e os estímulos eram sintetizados. O *corpus* era composto por 7 estímulos que variavam com relação ao VOT. Havia, então, entre /pa/ e /ba/, 5 variações. O experimento testou duas condições: (1) entre categorias (pares AB); e (2) dentro de categorias (pares AA e Aa).

Na tarefa de identificação, os informantes dividiram o contínuo em duas categorias. A divisão ocorre aproximadamente aos 30ms de VOT (da mesma maneira como apontaram Lisker & Abramson – 1967 *apud* Pisoni, 1974: p.287). A Ilustração 5 também mostra que os Informantes são mais lentos dentro das categorias (mais lentos no estímulo 4).

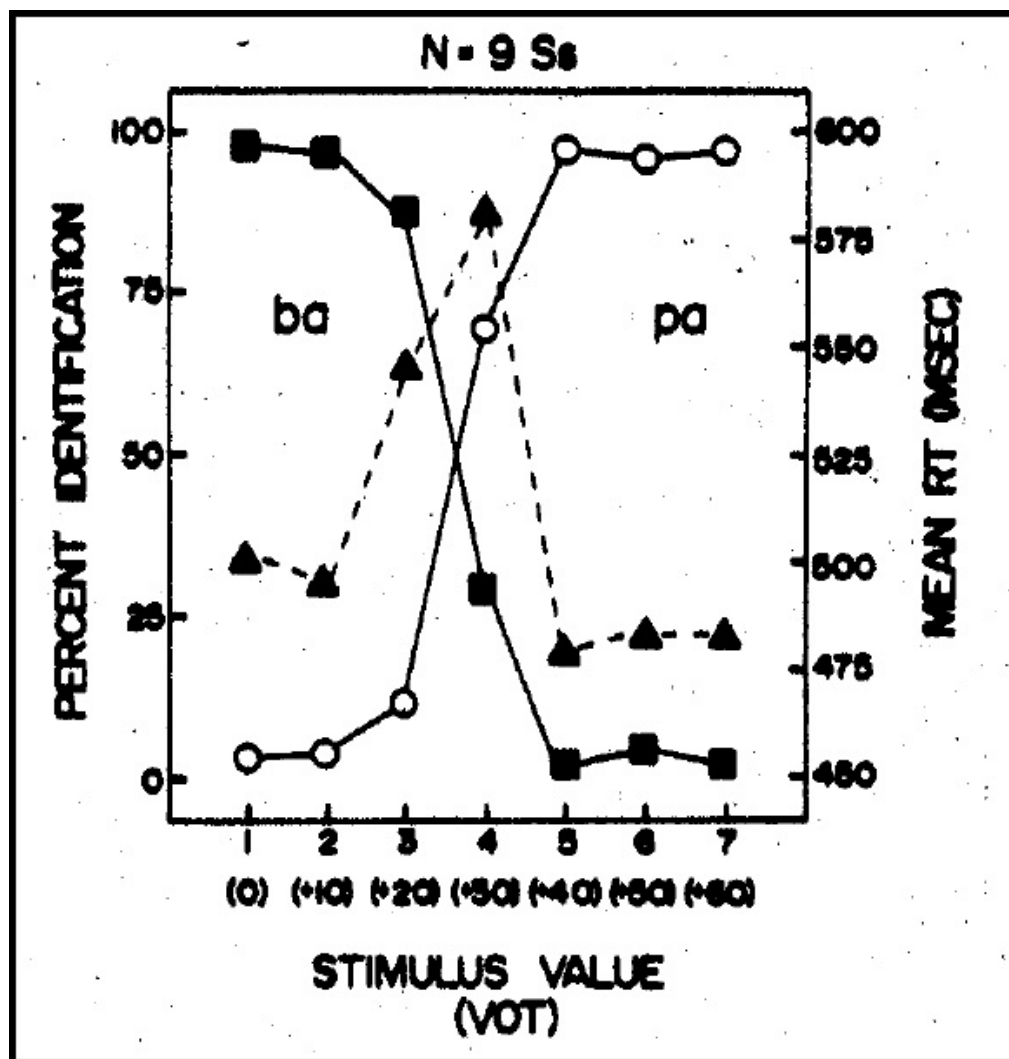


Ilustração 5: Resultados para a tarefa de identificação (cf. Pisoni, 1974: p.2881). É possível verificar que os informantes dividiram o contínuo em duas categorias. A divisão ocorre aproximadamente aos 30ms de VOT. As linhas com os quadrados preenchidos mostram o comportamento de [ba] e os círculos vazados, [pa]. Já a linha pontilhada com os triângulos correspondem às latências de identificação de cada resposta.

Já na tarefa em que os informantes têm que dizer se os estímulos são iguais ou diferentes por comparação, tem-se, para os resultados para respostas IGUAL, que os Informantes são significativamente mais rápidos para dados idênticos (AA) do que para os parcialmente iguais (Aa), i.e., eles conseguem utilizar informações acústicas de baixo nível mesmo os tendo categorizados como pertencentes a um mesmo segmento fonético. Isso mostra que os pares AA foram processados em estágio mais baixo e Aa em estágio mais avançado.

Diante desses dados, o autor passa a questionar se a percepção de consoantes é tão categórica quanto se dizia ser. A noção de percepção categórica decorre do *design* experimental adotado, de modo que os informantes são obrigados a trabalhar com algumas propriedades que fazem com que eles tenham que utilizar o código fonético ao invés dos níveis mais baixos da audição. Assim, o autor argumenta que os informantes não consideraram os estímulos Aa como sons da fala. No entanto, isso não é verdade, pois o nível de discriminação como igual é muito próximo. Isso mostra que eles consideraram Aa como som da fala e não como outro tipo de som.

Em suma, os resultados sugerem que as informações acústicas de nível baixo estão disponíveis até os níveis mais abstratos da representação fonética (até mesmo para consoantes). Isso decorre da inserção do tempo de reação que acabou revelando um estágio a mais de avaliação da percepção; um estágio além da categoria e da representação fonética. Segundo o autor, esses resultados podem mostrar como funcionam os mecanismos especiais para a percepção da fala e como ela interage com mecanismos mais gerais do processo de percepção.

1.2 Identificação por atributos secundários

Uma segunda linha de pesquisa sobre a percepção da fala é encontrada nos trabalhos de Cole & Scot (1974) e Stevens & Blumstein (1978)¹⁹. Nessas pesquisas, os autores apresentam argumentos para a proposta de que a fala é percebida com o auxílio de atributos secundários.

Cole & Scot (1974) argumentam que a tarefa da percepção da fala envolve a identificação de fonemas. Em suas palavras, segue a definição de fonema utilizada por eles:

¹⁹ Decidimos apresentar aqui apenas o trabalho de Cole & Scot (1974), uma vez que sua proposta parece ter sido “esquecida”, ao contrário do que ocorreu com a pesquisa de Stevens & Blumstein (1978). A título de exemplo de como o trabalho de Stevens & Blumstein (1978) não foi “esquecido”, mencionamos aqui dois “trabalhos de peso” que o endossa ou critica. Fowler (1986), para defender seu tratamento realista e direto para a percepção da fala comenta que a *teoria quântica* tinha um *insight* promissor sobre a relação entre o acústico e o articulatório. Contudo, seus primitivos anda eram estáticos. Em contrapartida, no trabalho de Albano (2001) há uma revisão e utilização da proposta de Stevens & Blumstein (1978) dentro de um arcabouço dinâmico.

“Por fonema, nós entendemos como um conjunto de fones (também chamados alofones) sobre os quais várias restrições são impostas, possivelmente incluindo invariância articulatória, mas não invariância acústica. Nosso uso do termo fonema corresponde ao de Liberman, Cooper, Shakweiler & Studdert-Kennedy (1967), que o definem como ‘o menor segmento que produz diferenças significativas entre sentenças [nota de rodapé 4, p. 431]’. Os termos ‘fonema’ e ‘segmento fonético’ são usados alternadamente neste trabalho. Como usado aqui, o fonema corresponde grosseiramente a ‘fone’ na fonologia gerativa – o segmento de uma língua mais próximo a sua realização física da sentença” (Cole & Scot, 1974: p.348²⁰).

Estudos iniciais com fala sintética (de Liberman e colegas) questionaram a noção de que havia um conjunto de características acústicas para cada fone produzido na fala. Assim, esses trabalhos lançaram olhar à transição presente tanto em contextos de vogal a vogal como em contexto de consoante a vogal. Eles propõem, então, que há uma transmissão em paralelo sobre o ponto de articulação de ambos os fonemas, i.e., a transição transmite ao mesmo tempo a informação sobre o ponto de articulação da consoante e da vogal (em caso de sílabas CV, por exemplo).

Stevens & Halle (1967: p. 90 *apud* Cole e Scot, 1974: p.349) definem o problema da seguinte forma:

“Nós, então, estamos diante de um problema: por um lado, a fala pode ser descrita de maneira satisfatória em termos de segmentos e traços, e, por outro, tais segmentos e traços não parecem estar diretamente presente em eventos observáveis na fala”²¹.

Esse problema sugere que as diferenças dos traços distintivos não são retiradas do sinal da fala, de modo que elas seriam derivadas de estruturas representacionais mais abstratas, tal como propõem Chomsky & Halle (1968).

²⁰ “By phoneme we mean a set of phones (so called allophones) upon which various constraints are imposed, possibly including articulatory invariance, but not acoustic invariance. Our use of the term phoneme corresponds to that of Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967), who define the phoneme as “the shortest segment that makes a significant difference between utterances [Footnote 4, p. 431].” The terms “phoneme” and “phonetic segment” are used interchangeably in this paper. As used here, the phoneme corresponds roughly to the “phone” in generative phonology — the segment of language closest to the physical realization of the utterance” (Cole & Scot, 1974: p. 348).

²¹ “We are thus faced with the problem that though on the one hand speech can be described in very satisfactory fashion in terms of segments and features, the latter do not seem to be directly present in the observable speech event” (Cole & Scot, 1974: p. 349).

Diante disso, os autores sugerem uma teoria que declara que há pelo menos três tipos de pistas envolvidas simultaneamente na percepção da fala: pistas invariantes; condicionadas pelo contexto; e promovidas pelo envelope da forma de onda.

Com relação às pistas invariantes, eles colocam que há invariantes acústicos que são responsáveis pela identificação de uma dada consoante em uma classe de fonemas, pares de fonemas e, em alguns casos, o fonema em particular. Tais pistas estão no domínio das frequências, i.e., no domínio acústico. As pistas condicionadas pelo contexto promovem a informação sobre a identidade de alguns fonemas. Elas seriam, basicamente, pistas de transição que, majoritariamente, carregariam informação sobre o ordenamento temporal dos segmentos dentro da sílaba. Também seriam responsáveis pela manutenção da integridade perceptual do sinal da fala, permitindo que a fala seja percebida como uma única cadeia [*stream*] perceptual. Por fim, as pistas dadas pelo envelope da forma de onda transmitem informações significativas sobre fonemas individuais e sobre as características prosódicas (acento e entonação) de uma sentença.

Em suma, as pistas invariantes e de transição estão integradas para formar a base perceptual de sílabas como unidades únicas, enquanto a informação de envelope é integrada às sílabas em unidades maiores. Assim, eles apresentam dados que endossam a sua proposta de que há 3 tipos de pistas que integram o mecanismo da percepção da fala. Passaremos, a seguir, a cada uma delas.

As pistas invariantes são definidas como pistas acústicas que acompanham um fonema em particular independentemente do contexto vocálico em que a consoante se insere. Por exemplo, para fricativas e africadas, há vários traços acústicos que os ouvintes usam para fazer a discriminação das suas produções, tais como duração, silêncio, aspereza²² do *onset* e vozeamento. Para tanto, os autores colocam os exemplos de diferenciação entre fricativas e fricativas e entre fricativas e africadas. Nas primeiras, há diferenças com relação ao pico espectral (a distribuição de energia do ruído). Com relação às africadas e fricativas, há diferenciação devido

²² A aspereza do *onset* é a forma abrupta pela qual as transições ocorrem quando se produz uma consoante oclusiva. A presença dessa aspereza diferenciaria fricativas de africadas, por exemplo. A diferenciação entre fricativas e africadas (cf. Sandra Madureira, em comunicação pessoal) pode ser realizada pelo *rise time* do sinal acústico.

à presença de silêncio nas africadas.

As pistas invariantes também estão presentes quando há transições em sílabas CV. Por exemplo, /f, v, θ, ð, m, n/ requerem identificação simultânea das pistas invariantes e de transição. As fricativas /f, θ/ compartilham dos mesmos invariantes acústicos (ruído de alta frequência de baixa amplitude). Já as nasais são caracterizadas por uma única barra de energia de ressonância de frequência baixa e nenhum ruído de alta frequência. Essas características diferenciam /f/ de /θ/, /v/ de /ð/ e /m/ de /n/ como pares de fonemas diferentes²³. Ainda mais, para diferenciar entre cada um desses fonemas, é preciso levar em conta as transições das vogais, uma vez que as consoantes anteriores (diante da vogal /a/) têm um F2 ascendente e as mais centrais, F2 descente.

Ainda com referência às pistas invariantes das consoantes oclusivas, tem-se como característica principal o silêncio, que diferencia as oclusivas de outras consoantes. Em monossílabos CV do inglês, ficam evidentes todas as pistas acústicas das oclusivas: *burst*; seguido de um pequeno período de abertura (menor para /b/, intermediário para /d/; e maior para /g/). As surdas não apresentam essas aberturas, pois há aspiração.

Por sua vez, ao tratar do vozeamento, Abramsom & Lisker (1964 *apud* Cole & Scot, 1974: p.355) apontam que o VOT é mais longo para surdas e mais curto para sonoras. Os valores de VOT também aumentam à medida que o ponto de articulação se posterioriza no trato vocal. O VOT, portanto, é um bom parâmetro para distinguir consoantes surdas de sonoras. Eimas, Siqueland, Jusczyk & Vigorito (1971 *apud* Cole & Scot, 1974: p.356) demonstraram com experimentos de percepção que recém nascidos respondem categoricamente a esses estímulos, sugerindo que o vozeamento é uma propriedade inata do sistema perceptual humano.

No entanto, esses parâmetros variam de acordo com o contexto: em consoantes sem soltura em coda silábica, a diferença de vozeamento é garantida pelo tamanho da vogal; em posição medial, o tamanho da oclusão é suficiente para

²³ O leitor pode estar pensando que em realidade essas características (as transições) diferenciariam /f/ de /v/ ou invés de /f/ de /θ/. No entanto, /f/ e /v/ são diferenciados pela sonoridade e não pelas transições.

opor uma surda a uma sonora. (cf. Lisker, 1957 *apud* Cole & Scot, 1974: p.356). Assim, o modo de articulação de uma consoante oclusiva é sinalizado por um traço acústico particular (VOT, tempo de oclusão, duração da vogal ou soltura da consoante) em um dado contexto específico.

Os autores propõem que, embora não haja uma relação biunívoca entre o sinal acústico e articulação, F2 é o principal correlato acústico para ponto de articulação. Delattre, Liberman & Cooper (1955), na tentativa de encontrar um *locus* para cada consoante, argumentaram que as transições de F2 representam diretamente os movimentos articulatórios do ponto de produção de uma consoante para a posição da vogal seguinte. Para eles, era possível encontrar um *locus* fixo para cada consoante, pois o ponto de articulação de cada uma delas seria em grande parte fixo. Diante disso, Cole & Scot tomam a frequência do *burst* e de F2 como bons parâmetros para diferenciar os pontos de articulação de consoantes.

Os autores ainda citam que Fischer-Jørgensen (1972 *apud* Cole & Scot, 1974: p.358) verificaram que todos os parâmetros são importantes para a percepção (*burst*, transição de vogais e aspiração) das oclusivas do dinamarquês. Ainda mais, cada parâmetro é importante em um contexto diferente:

“A importância das explosões iniciais antes de /i/ e /u/ encontradas na presente investigação também são de interesse para uma teoria geral da percepção da fala. Em várias ocasiões (e.g., 1970), Liberman tem enfatizado que a fala é caracterizada pela transmissão em paralelo e pela sobreposição de pistas (desde que as transições tenham informações sobre as consoantes e vogais) e que o traço característico da percepção da fala é então o de decodificar tais sobreposições de pistas. Ele faz uma exceção para fricativas (no dinamarquês, a troca de /f/ e /s/ antes de /i/, /a/ e /u/ também mostra que o ruído fricativo é decisivo aqui), mas ele poderia ter feito uma exceção para combinações de oclusivas e vogais também. Os exemplos mais frequentemente quotados por Liberman são /bæg/ e as sílabas /di/ e /du/. /bæg/ é de fato um bom exemplo, mas é notável que a transmissão de /di/ em dinamarquês falado parece ser de nenhuma importância para a percepção. Não pode ser negado que as pistas de sobreposição são muito frequentemente encontradas na fala, e isso mantém o mérito do grupo do *Haskins* ter demonstrado esse fato, mas seu papel na percepção da fala natural pode ter sido um pouco subestimada” (Fischer-Jørgensen 1972: p.164 *apud* Cole

& Scot, 1974: p.358-359²⁴).

Diante desses dados, Cole & Scot vão dizer que experimentos muito controlados ocultam o fato de as consoantes possuírem um conjunto de traços que as identificam, pois há, em realidade, um conjunto invariante de traços que, na ausência de um deles, um outro se torna responsável pela sua percepção, i.e., “a percepção de consoantes oclusivas claramente envolve a integração de pistas tanto invariantes quanto condicionadas pelo contexto” (Cole & Scot, 1974: p.359). Assim, eles decidem realizar um experimento para tentar testar a sua hipótese.

Nesse experimento, o *burst* do *onset* da consoante foi mantido fixo para cada consoante e toda a energia entre o *burst* e o início da vogal foi removido. Em seguida, as vogais das sílabas foram “coladas” trocadas em cada consoante, i.e., passou-se a ter sílabas com um /b/ com transições de /i/ mas com uma vogal /u/ no lugar, por exemplo. Nas ilustrações 6 e 7, há todos os estímulos produzidos para o experimento.

Os resultados encontrados mostram que a percepção “correta” apenas não ocorre para as sílabas com /g/. Nos outros casos, os informantes buscaram a pista invariante presente no *burst*. Por sua vez, quando não há soltura, em decorrência de posição medial ou em sílabas finais não acentuadas, a identificação do ponto de articulação da consoante parece depender muito mais das transições das vogais adjacentes.

²⁴ “The importance of the initial explosions before /i/ and /u/ found in the present investigation has also some interest for the general theory of speech perception. On several occasions (e.g., 1970) Liberman has emphasized that speech is characterized by parallel transmission and overlapping cues (since the transitions contain information about both consonants and vowels) and that the characteristic feature of speech perception is therefore the decoding of such overlapping cues. He makes an exception for fricatives (interchange of Danish /ʃ/ and /s/ before /i/, /a/ and /u/ also shows that the fricative noise is alone decisive here), but he might have made an exception for some combinations of stops and vowels too. The examples most frequently quoted by Liberman are /baeg/ and the syllables /di/ and /du/. /bag/ is in fact a good example, but it is remarkable that the transition in spoken Danish /di/ seems to be of no importance for perception. It cannot be denied that overlapping cues are very often found in speech, and it remains the merit of the Haskins group to have demonstrated this fact, but its role in the perception of natural speech may have been somewhat overemphasized” (Cole & Scot, 1974: p.358-359).

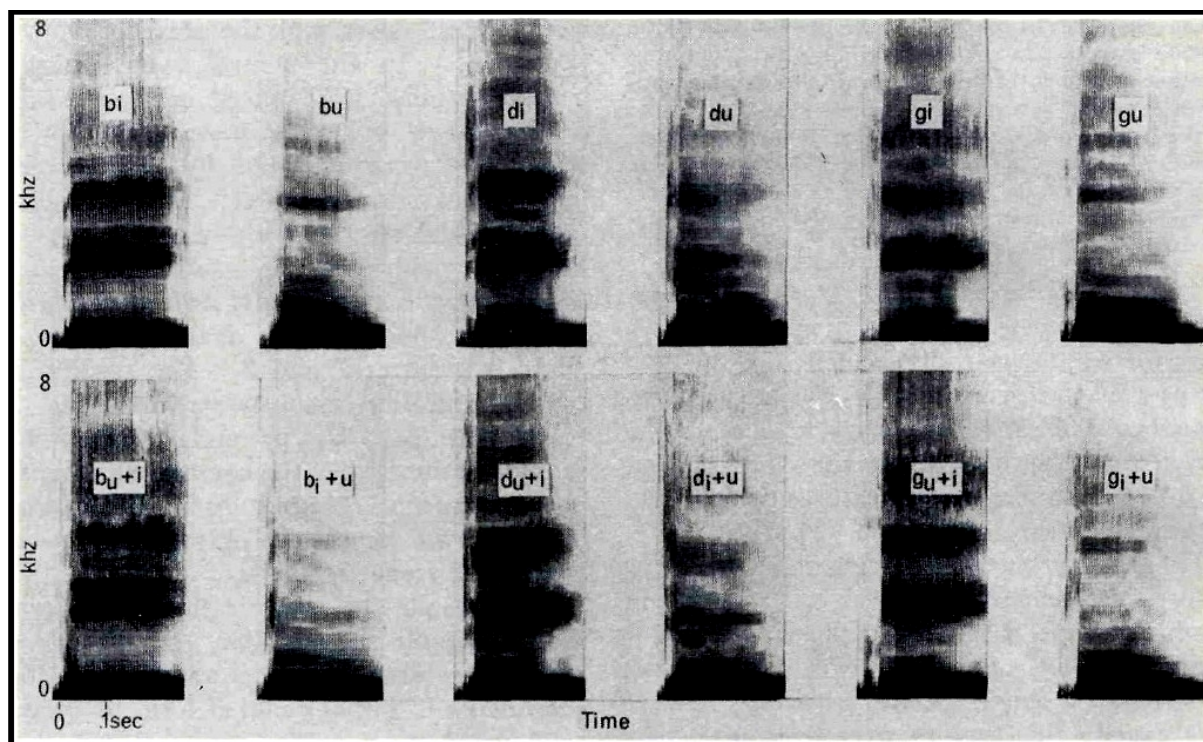


Ilustração 6: Espectrogramas dos estímulos utilizados no experimento de Cole & Scot (1974: p.360): consoantes sonoras.

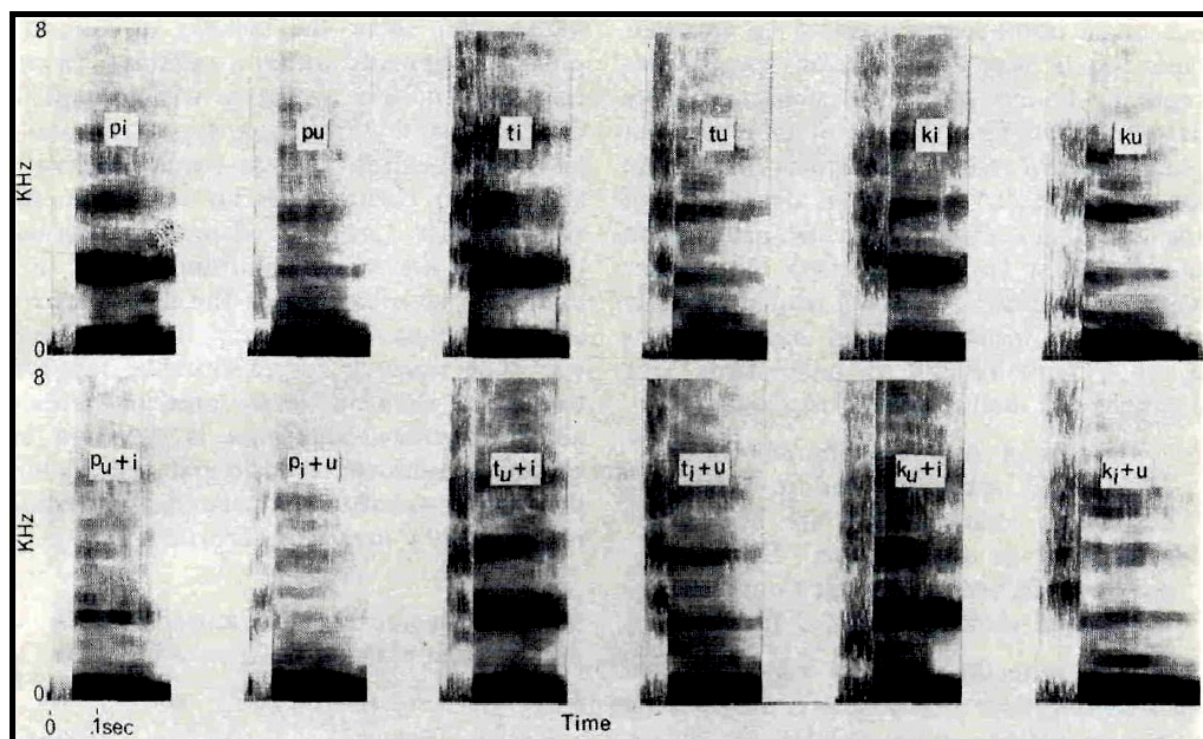


Ilustração 7: Espectrogramas dos estímulos utilizados no experimento de Cole & Scot (1974: p.361): consoantes surdas.

Em suma, as pistas invariantes apresentam contribuições importantes para a percepção da fala. De acordo com a pesquisa realizada pelos autores, a informação sobre os fonemas individuais é fornecida por pistas de transição condicionadas por contexto, enquanto a maior fonte de formação é provida pelos traços acústicos invariantes. Eles ainda citam que em experimentos de reconhecimento de fala, quando se alteram as transições, a porcentagem de acerto entre as duas formas (normal e alterada) é muito próxima (29,3 e 27,8, respectivamente). Quando eles alteram as pistas invariantes, a porcentagem de erro cai para 12,7. Assim,

“traços invariantes para fonemas consonantais são necessários para uma percepção acurada da fala e pistas invariantes de consoantes por si só provêm informação suficiente para permitir que os ouvintes compreendam sentenças isoladas. Parece que a maior fonte de informação sobre a identidade de fonemas consonantais durante a percepção da fala é fornecida por pistas invariantes” (Cole & Scot, 1974: p.262²⁵).

Tendo apresentado seus argumentos sobre a importância e sobre a existência de pistas invariantes, os autores passam a considerar a importância de se ter transições na produção da fala. Basicamente, eles vão argumentar que elas ajudam a manter um ordenamento temporal da fala.

Os autores mostram que (cf. Warren, Obusek, Farmer & Warren, 1969 *apud* Cole & Scot, 1974: p.362) as pessoas conseguem perceber individualmente os sons (mesmo quando os experimentadores inserem *hisses*, vogais, *buzzes* e tons de 200ms junto aos sons), mas não conseguem identificar seu ordenamento. Em contrapartida, quando vogais são colocadas em sequência, com intervalos de 50ms e com *onset* e *offset* graduais, a sua identificação de ordenamento temporal passa a ser perfeita (cf. Warren & Warren, 1970 *apud* Cole & Scot, 1974: p.362). Tais dados são discrepantes com relação ao fato de a fala ser claramente percebida na faixa de 40 a 50 fonemas por segundo em fala rápida e, na normal, com fonemas de pelo menos 80 milissegundos (cf. Foulke & Sticht, 1969; Efron, 1963 *apud* Cole & Scot, 1974: p.363).

A importância da existência de transições também foi salientada por Liberman

²⁵ “Thus, invariant features for consonant phonemes are *necessary* for accurate perception of speech, and the invariant consonant cues by themselves provide sufficient information to allow listeners to comprehend isolated sentences. It appears that the major source of information about the identity of consonant phonemes during speech perception is provided by invariant cues” (Cole & Scot, 1974: p.262).

e colegas (1967). Eles sugerem que a fala é um código especial e que as transições são importantes para garantir seu ordenamento temporal. Por exemplo, /da/ e /ad/ apresentam transições espelhadas e isso seria uma pista para perceber onde está a consoante na sílaba.

Bregman & Campbel (1971 *apud* Cole & Scot, 1974: p.363) realizaram um experimento em que era apresentado aos informantes um *looping* de tons altos e baixos. Os informantes não conseguiram ordenar as ocorrências de modo que ouviam duas cadeias simultâneas (uma de tons altos e outra de tons baixos). Os autores sugeriram que isso é uma tendência do sistema auditivo primário de agrupar sons semelhantes em canais diferentes. Assim, Cole & Scot (1974: p.363-364²⁶) perguntam:

“Se o sistema auditivo agrupa estímulos em cadeias separadas baseadas em características de frequências, por que a fala falha em segregar em cadeias distintas, uma composta por fricativas e outra por *bursts* de frequências altas, a outra composta por energia de ressonância de nasais, vogais e semivogais?”

A hipótese é que as transições servem para evitar uma cadeia auditiva de discriminação. Bregman & Dannenhring (1973 *apud* Cole & Scot, 1974: p.364) mostraram que o efeito diminui quando se insere *glides* entre os tons. Em um experimento novo, eles mostram que, se a porção invariante da sílaba é retirada, os informantes percebem a consoante oclusiva “homorgânica” da sílaba (ex: /ma/ e /fa/ são percebidos como /ba/ e /na/ e /sa/, como /da/). i.e., mantém-se a transição (ou a informação sobre o ponto de articulação da consoante) e a vogal. A hipótese deles, nesse experimento, é a de que se uma sílaba for repetida num *looping* aos informantes, espera-se que eles discriminem o sinal, separando-o em duas cadeias: uma com o ruído invariante; e outra com a sílaba percebida como oclusiva mais vogal (que seria a transição mais o estado estacionário da vogal). Isso ocorreu! Numa sílaba como /sa/, os informantes ouviram um ruído mais a sílaba /da/ e em /fa/, um “ruído de gato” mais /ba/.

Em outro experimento, eles eliminaram as transições e “colocaram” o início no

²⁶ “If the auditory system groups stimuli into separate streams based on frequency characteristics, why does speech fail to segregate into different streams, one composed of fricatives and other high-frequency bursts, the other composed of resonant energy of nasals, vowels, and semivowels?” (Cole & Scot, 1974: p. 363-364).

estado estacionário da vogal. Isoladamente, isso não altera a percepção da sílaba. Mas quando os estímulos foram inseridos num *looping*, os informantes não ouviram a consoante oclusiva como uma ilusão, mas sim a vogal. Ainda mais, essa segregação ocorreu após 3 ou 4 repetições. Enquanto em sílabas normais, isso demorou em média 60 vezes. Diante disso, os autores salientam que as transições parecem inibir a segregação do sinal.

Por fim, a terceira pista para a percepção da fala se refere ao envelope da forma de onda. Os autores colocam que, como o domínio da amplitude e do tempo têm sido considerados como secundários na percepção da fala, levanta-se a hipótese de que “o envelope da forma de onda é extraído como um todo e é um tipo de pista discreta no mesmo sentido de que pistas invariantes e de transição são discretas.” (Cole & Scot, 1974: p.367).

Com relação aos traços prosódicos, eles apresentam três argumentos: (i) o acento e a entonação são influenciados pela amplitude e tempo (separadamente) com relação à frequência, i.e., a amplitude e a duração dos segmentos auxiliam na percepção de sílabas tônicas e ajudam a gerar leituras diferentes em sentenças “ambíguas” (cf. Scholes, 1971 e Fry, 1955 *apud* Cole & Scot, 1974: p.367); (ii) o período de silêncio entre as sílabas promove a segmentação diferente de sentenças (*lighthouse keeper* ou *light housekeeper*); e (iii) Liberman (1960 *apud* Cole & Scot, 1974: p.368) mostrou que o *pitch*, a amplitude e a duração também auxiliam na percepção do acento.

Já com relação aos traços fonêmicos, eles mostram exemplos da relação entre informação fonética e o envelope da forma de onda como independentes da frequência: “esses exemplos demonstram que a informação temporal transmitida pelo envelope da forma de onda é frequentemente crítica para realizar distinções fonêmicas, mesmo com total ausência de informação de frequência” (p.368). Por exemplo, se se estender o período de silêncio da oclusão de /b/ em *rabid*, tem-se *rapid* (cf. Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Lisker, 1957 *apud* Cole & Scot, 1974: p.368); se se inserir silêncio entre /s/ e /l/ em *slit*, tem-se como consequência *split*. (cf. Bastian, Eimas & Liberman, 1961 *apud* Cole & Scot, 1974: p.368); e se se aumentar o silêncio entre as sílabas de *topic* causaremos a sensação de dois /p/ (*top pic*) – cf. Pickett & Decker, 1960 *apud* Cole & Scot, 1974:

p.368).

Diante desses dados, os autores concluem que os experimentos têm mostrado que os aspectos temporais (negligenciados nas teorias de percepção) conseguem traçar características perceptuais da fala tanto no âmbito prosódico como no segmental. Tais características (amplitude, duração e *pitch*) podem ser extraídas do envelope de forma de onda.

Em suma, os autores propõem a existência de três fontes de pistas perceptuais (*burst* e frequências de estados estacionários, pistas de transição e pistas de envelope). Essas três fontes não implicam a existência de três processadores diferentes, embora haja evidência para sustentar a existência de três processadores (ou mecanismos).

Os dados sobre a segregação da sílaba /sa/ em repetição (ou em *looping*) mostram que há o processamento da vogal com sua transição de uma maneira diferente ao invariante (só o /s/, ou ruído). “Este achado sugere que a porção vocálica de uma sílaba que possui transições é processada como uma unidade da sílaba enquanto a porção do ruído invariante é processado separadamente” (Cole & Scot, 1974: p.370). Soma-se a isso o fato de as vogais sempre serem ouvidas como fala mesmo quando a segregação ocorre, enquanto a porção invariante é percebida como ruído ou ambiência.

Assim, a existência de três mecanismos integrados de análise acústica é necessária para gerar o máximo de redundância. Segundo os autores, as vantagens de se ter um modelo com três processadores são: (i) se o modelo só “olhasse” para os invariantes haveria dois problemas: 1) em muitos casos não há correspondência de um para um entre o sinal acústico e um fonema; e 2) precisaríamos processar os fonemas de uma maneira muito rápida se fôssemos processar um fonema por vez (independentemente); (ii) a possibilidade de pensar a codificação da fala a partir das pistas de transição (em unidades do tamanho da sílaba, tal como sugerem Liberman e colegas, 1967) colaboraria para a economia de tempo do processamento. No entanto, só as pistas de transição não são suficientes para codificar o sinal acústico satisfatoriamente; (iii) a integração das pistas invariantes e as pistas de transição gerariam um volume de informação satisfatório. No entanto, há a amplitude e a

duração que desempenham papéis importantes na percepção da fala. O envelope da forma de onda transmite informações de maneira mais lenta que a forma de onda e isso pode ser uma vantagem para um processador “limitado”. A vantagem de se ter um processador do envelope é similar ao fato de se ter um processador de transições: o envelope sofre da redução de informação fonética específica, mas beneficia a transmissão de informação sobre um segmento maior na fala corrida; (iv) a integração dos processadores é essencial, uma vez que

“o modelo com três processadores demanda que cada processo seja integrado com outro da mesma maneira que os mantém independentes. A integração é essencial uma vez que sabemos que alguns julgamentos fonéticos podem ser feitos apenas na presença de pistas de transição e de estado estacionário ou de pistas de envelope (Cole & Scot, 1974: p.371²⁷)”;

(v) a integração dos processadores implica em adotar a sílaba como unidade perceptual mínima. A diferença desta abordagem é que é preciso considerar todos os parâmetros ao invés de apenas as transições (tal como Liberman e colegas, 1967). Assim, o núcleo silábico ajudaria a dar a informação sobre a qual invariante pertence a consoante (uma vez que ela indica de onde veio a constrição do trato vocal). Isso também implica no fato de se considerar que o núcleo vocálico é melhor codificado diretamente das formas de onda do que de uma decodificação de fonemas feita pelo sistema motor; (vi) as pistas de envelope ajudam as outras duas pistas a identificar fronteiras de sílabas. Talvez seu papel mais importante seja o de integrar diversas sílabas em sentenças e, assim, extrair os traços prosódicos.

1.3 Percepção dos sons por outros animais

Uma terceira linha de pesquisa sobre a percepção da fala de orientação auditiva é encontrada nos trabalhos de Kuhl & Miller (1975), Kuhl (1981) e Lotto, Kluender & Holt (1997)²⁸. Kuhl & Miller argumentam que tanto a análise quanto as

²⁷ ““The three-process model demands that each process be integrated with the others as well as remaining independent. Integration is essential since we know that some phonemic judgments can only be made in the presence of steady-state and transitional cues, or steady-state and envelope cues” (Cole & Scot, 1974: p.371).

²⁸ Tal como nas seções anteriores, apenas apresentaremos o trabalho de Kuhl & Miller (1975). A escolha da sua apresentação decorre da sua importância nos estudos fônicos, uma vez que se trata de um marco sobre a percepção dos sons da fala por outros animais.

técnicas de síntese de fala não conseguiram apontar satisfatoriamente para os elementos invariantes que compõem a nossa percepção. A “falta” de relação biunívoca entre as características acústicas e articulatórias fizeram surgir respostas “especiais” para a percepção da fala. Grosso modo, são duas: (i) algumas classes de sons da fala são reconhecidas por “processos especiais”, i.e., especula-se a existência de “detectores de traços fonéticos”; e (ii) a Teoria Motora (ver seção 2 do Capítulo 2) sugere que há um conhecimento tácito sobre os resultados acústicos de manobras articulatórias que faz parte da percepção da fala. Ambas as propostas sugerem que a percepção da fala é uma habilidade específica da espécie humana:

“Tal como colocou Liberman, infelizmente, nada é conhecido sobre a maneira como animais não humanos percebem a fala [...] entretanto, nós devemos supor que, na ausência de um decodificador de sons da fala, os animais não perceberiam a fala tal como nós a fazemos, até mesmo num nível fonético” (Liberman, 1970 apud Kuhl & Miller, 1975: p.69²⁹).

Diante disso, os autores se perguntam se as chinchilas (devido às semelhanças do seu aparelho auditivo com o humano) conseguiriam classificar corretamente uma quantia grande de sílabas naturalmente pronunciadas com relação ao contraste de sonoridade. Para isso, eles decidem realizar dois experimentos.

No primeiro experimento, 4 chinchilas foram treinadas para perceberem a diferença entre /t/ e /d/ em sílabas CV. Depois de treinadas, elas foram capazes de identificar novas instâncias de /t/ e /d/, incluindo sílabas produzidas por novos falantes, em novos contextos vocálicos e em fala sintetizada.

No segundo experimento, estímulos sintetizados que variavam o VOT entre /ta/ e /da/ foram apresentados às chinchilas treinadas, às não treinadas e a humanos. Tanto a maneira de rotular os estímulos quanto o estabelecimento de fronteiras fonéticas foram similares a todos os animais do experimento e os humanos.

Os resultados desses experimentos demonstram que a distinção de sonoridade em consoantes em posição inicial não é uma habilidade exclusiva de

²⁹ “As Liberman stated, 'Unfortunately, nothing is known about the way non-human animals perceive speech ... however, we should suppose that, lacking the speech-sound decoder, animals would not perceive speech as we do, even at phonetic level'” (Kuhl & Miller, 1975: p. 69).

humanos. Tais distinções (tanto na fala natural, que apresenta os traços críticos misturados a outros, quanto na fala sintetizada, que dispõe dos traços distintivos isolados) podem ser realizadas por outros mamíferos, “sem a mediação de dispositivos fonéticos” (Kuhl & Miller, 1975: p.71).

Em outras palavras, esses resultados apontam para o fato de que tais distinções realizadas pela percepção da fala não é um “privilegio” de humanos, de modo que essas questões merecem explicações futuras. Para os autores,

“o fato de que as chinchilas respondem à fala sintética, como se uma mudança qualitativa abrupta ocorresse na pequena região de vozeamento no *continuum* do VOT precisamente no ponto onde muitas línguas separam duas categorias fonêmicas, apoia a ideia de que as oposições dos sons da fala foram selecionadas para serem altamente distintivas ao sistema auditivo. (Kuhl & Miller, 1975: p. 72³⁰)

Kuhl & Miller (1975) terminam seu texto sugerindo que mais experimentos com animais devem ser realizados para que se tenha mais dados para apoiar ou refutar a necessidade de “processamento especial” para a fala³¹.

1.4 Sumário da seção

Nesta seção, consideramos a percepção da fala a partir das pesquisas que tomaram estritamente seus aspectos acústicos como objetos fundamentais para a sua percepção. Iniciamos por abordagens que a consideraram em uma interface com experimentos relacionados à memória (Pisoni, 1973, 1974). Passamos por experimentos que chegaram a propor vários processadores para extrair as informações presentes em níveis secundários (Cole & Scot, 1974). Por fim, abordamos os achados sobre a percepção da fala por animais (Kuhl & Miller, 1975).

Especificamente, o trabalho de Kuhl & Miller (1975) problematiza o fato da fala ser considerada “especial”, pois eles tentam mostrar que outros animais também

³⁰ “The fact that the chinchilas respond to the synthetic speech as though an abrupt qualitative change occurs in the short voicing-lag region of the VOT continuum at precisely the place where many languages separate two phonemic categories lends, support to the idea that speech-sound oppositions were selected to be highly distinctive to the auditory system” (Kuhl & Miller, 1975: p. 72).

³¹ Aos que se interessarem pelas questões acerca da percepção dos sons da fala por animais e humanos consultar: KUHLE (1991), KUHLE (1993) e KUHLE & IVERSON (1995).

conseguem fazer as mesmas distinções que os humanos realizam ao perceber a fala. A proposta acerca do caráter especial da fala é um tópico da Teoria Motora da Percepção da Fala. Trataremos desse tópico com mais atenção na seção 2 do segundo capítulo. Nas seções que se seguem, apresentaremos as abordagens que concebem os primitivos perceptuais como sendo de base articulatória.

2 COMO ARTICULAÇÃO

Nesta seção, apresentaremos as teorias que tomam objetos articulatórios como primitivos da percepção da fala. Primeiramente, trataremos da Teoria Motora da Percepção da Fala (Mattingly & Liberman, 1985) e, em seguida, da Teoria do Realismo Direto da Percepção da Fala (Fowler, 1996). Será possível notar que, embora as duas teorias tomem a articulação como primitivos de análise, elas apresentam diferenças significativas no modo de encarar a fala, sendo a primeira de natureza idealista e a segunda, realista.

2.1 Como gestos pretendidos

Muitos de nós (se não todos), ao ver uma luta de boxe na televisão, já deve ter se perguntado por qual motivo um dos lutadores ganhou a luta. Nas lutas que acabam por *knockout*, é fácil entender o que levou um dos lutadores à vitória: um deles disparou ou um único golpe certo ou um conjunto de ataques que derrubaram o outro competidor. No entanto, quando a luta segue *round* após *round* até o fim e você se depara com o juiz levantando o braço do vencedor no final, você se pergunta por que este competidor e não o outro ganhou, sendo que os dois bateram e apanharam durante os 10 *rounds*? A resposta é técnica. Só o juiz, um especialista, sabe interpretar qual deles fez uma luta melhor. Isto é, só ele sabe quem pretendia dar os golpes que poderiam ter decidido a luta antes. Em suma, o resultado final deste caso é decidido por quem se arriscou mais com a intenção de encerrar a luta por *knockout*.

Na percepção da fala, não é diferente. Na teoria que apresentaremos em seguida, os primitivos de análise são os gestos da fala. Contudo, não se trata dos gestos que realmente ocorreram (os que “provocam *knockout*”), e sim os movimentos articulatórios pretendidos. Aqui, os ouvintes são os especialistas que conseguem identificar quais são as intenções dos gestos. Passaremos, então, à proposta da Teoria Motora da Percepção da Fala – TM – (Mattingly & Liberman, 1985).

2.1.1 Teoria Motora da Percepção da Fala

A TM tem suas origens quando Liberman e colegas, nos Laboratórios *Haskins*, buscavam quais eram os padrões acústicos responsáveis pela percepção de um som³². Eles encontraram que os padrões acústicos de fala sintetizada tinham que ser alterados para que um “percept” fonético invariante fosse percebido como tal em diferentes contextos. Esses dados indicavam que os objetos da percepção não estavam no sinal acústico. Tais objetos talvez estivessem em processos motores subjacentes. Com isso, após aproximadamente 30 anos desses achados, os autores estabelecem quais são as bases para uma teoria de percepção da fala de base articulatória.

O primeiro ponto estabelecido por eles se refere aos objetos da percepção, que seriam gestos fonéticos pretendidos pelo falante, “representados no cérebro como comandos motores invariáveis que ‘chamam’ os movimentos dos articuladores em certas configurações linguisticamente significativas” (Liberman & Mattingly, 1985: p.237).

Tais comandos seriam a realidade física subjacente às noções fonéticas tradicionais, i.e., trata-se de um olhar mais minucioso sobre o que estaria por trás da realização fonética de um som. Desta forma, os segmentos fonéticos “são grupos simples de um ou mais desses eventos elementares; assim, [b] consiste de um gesto de oclusão labial e [m] de um mesmo gesto combinado com um gesto de abaixamento de véu” (Liberman & Mattingly, 1985: p.238).

Em contrapartida,

“fonologicamente, é claro, os gestos por si só devem ser vistos como grupos de traços, tais como ‘labial’, ‘oclusivo’, ‘nasal’, mas esses traços são atributos dos eventos gestuais, não eventos em si. Perceber uma sentença, então, é perceber um padrão específico de gestos pretendidos” (Liberman & Mattingly, 1985: 238³³).

³² Um maior detalhamento de como se deu esse processo se encontra na seção 2 do Capítulo 2 deste trabalho.

³³ “Phonologically, of course, the gestures themselves must be viewed as groups of features, such as ‘labial’, ‘stop’, ‘nasal’, but these features are attributes of the gestural events, not events as such. To perceive an utterance, then, is to perceive a specific pattern of intended gestures” (Liberman & Mattingly, 1985: p. 238).

Em suma, a noção de gestos pretendidos surge devido aos efeitos de coarticulação e pelo fato de os gestos não serem manifestados diretamente no sinal acústico ou em movimentos articulatórios observáveis.

O segundo ponto da teoria (corolário do primeiro) coloca a produção e a percepção da fala intimamente ligadas, uma vez que elas compartilham de um mesmo conjunto de invariantes.

O *link* entre a produção e a percepção não é uma associação aprendida, ela é inatamente especificada, requerendo “apenas desenvolvimento epigenético para trazê-la à tona” (Liberman & Mattingly, 1985: p.238). Diante disso, os autores postulam a especialização da fala:

“Neste ponto, a percepção dos gestos ocorre em um módulo especializado, importantemente diferente do módulo auditivo, responsável também pela produção de estruturas fonéticas, e parte da grande especialização da linguagem” (Liberman & Mattingly, 1985: p.238³⁴).

Os autores argumentam que a existência de um módulo especializado para a linguagem teria uma função adaptativa, uma vez que o módulo tem que

“fazer a conversão do sinal acústico para o gesto automaticamente e, então, fazer os ouvintes perceberem as estruturas fonéticas sem mediação (ou tradução) das aparições acústicas que os sons, em níveis puramente psicoacústicos, eram esperados de ter” (Liberman & Mattingly, 1985: p.238³⁵).

Tais propostas mostram que “a teoria não é logicamente sem significado, tampouco biologicamente inimaginável” (Liberman & Mattingly, 1985: p.238). Isso faz com que a teoria ganhe maior poder explicativo que as abordagens que assumem primitivos de natureza auditiva. Segundo os autores, toda teoria de percepção da fala vai se deparar com o fato de que “os padrões acústicos proximais devem ser os objetos distais percebidos. Embora a relação entre o gesto e o sinal acústico não seja estreita” (Liberman & Mattingly, 1985: p.238).

³⁴ “On this claim, perception of the gestures occurs in a specialised mode, different in important ways from auditory mode, responsible also for the production of phonetic structures, and part of the larger specialization for language” (Liberman & Mattingly, 1985: p.238).

³⁵ “The adaptative function of the perceptual side of this mode, the side which the motor theory is directly concerned, is to make the conversion from acoustic signal to gesture automatically, and so to let listeners perceive phonetic structures without mediation by (or translation from) the auditory appearances that the sounds might, on purely psychoacoustic grounds, be expected to have” (Liberman & Mattingly, 1985: p.238).

Ao assumir os gestos pretendidos como primitivos da percepção é possível propor que o *timing* dos gestos realizados não é o mesmo dos gestos envolvidos em um dado símbolo, pois

“os movimentos dos gestos em decorrência de um único símbolo não são tipicamente simultâneos, e os movimentos decorrentes de símbolos sucessivos sempre se sobrepõem extensivamente. Esta coarticulação significa que a mudança do formato do trato vocal, e, por consequência, o sinal resultante, é influenciado por vários gestos ao mesmo tempo. Portanto, a relação entre gesto e sinal é sistemática de uma maneira que é peculiar à fala” (Liberman & Mattingly, 1985: p.238-239³⁶).

Devido a isso, um mesmo gesto pode possuir mais de um correspondente acústico. Basicamente, as transições variam de acordo com o contexto (vogal adjacente à consoante) e o fato de que percebemos os gestos pretendidos (e não os realizados) faria com que ainda assim percebamos um dado som (com características acústicas diferentes) como sendo a mesma consoante.

Como podemos notar, diferentemente das teorias de base auditiva elencadas nas seções anteriores, a TM apresenta um tratamento diferente para a percepção da fala, uma vez que

“[...] a percepção da fala não é explicada por princípios que se aplicam à percepção de sons em geral, mas deve, ao contrário, ser vista como uma especialização para os gestos fonéticos. A incorporação de uma conexão biologicamente baseada entre a percepção e a produção, esta especialização impede os ouvintes de tomar o sinal como sons ordinários, mas permite que eles usem a relação sistemática, ainda especial, entre sinal e gesto para perceber o gesto. A relação é sistemática porque resulta de uma dependência de fidelidade entre os gestos, movimentos articulatórios, formatos de cavidade oral e sinal acústico. Ela é especial porque ocorre apenas na fala” (Liberman & Mattingly, 1985: p.240-241).

Diante do fato de a fala receber estatuto especial na teoria, o ouvinte consegue utilizar as variações sistemáticas das transições como uma informação sobre a coarticulação de uma consoante invariável com diversas vogais para

³⁶ “[...] the movements for gestures implied by a single symbol are typically not simultaneous, and the movements implied by successive symbols often overlap extensively. This coarticulation means that the changing shape of the vocal tract, and hence the resulting signal, is influenced by several gestures at the same time. Thus, the relation between gesture and signal, though certainly systematic, is systematic in a way that is peculiar to speech” (Liberman & Mattingly, 1985: p.238-239).

perceber seu gesto. Assim, a percepção “não requer associação arbitrária correspondente ao estágio auditivo” para lhe conferir um rótulo fonético. Aqui, a categoria fonética nomeia a si mesmos (Studdert-Kennedy, 1976).

Os autores ainda propõem que a TM merece o epíteto de motora duas vezes, pois (i) o próprio objeto da percepção fonética é um evento motor; e (ii) as adaptações do sistema motor para controlar os órgãos do trato vocal tomam precedência na evolução da fala. Isto é, ao permitir que os gestos fonéticos realizem de maneira coarticulada uma produção rápida, o sistema perceptual nota suas consequências acústicas complexas se desenvolverem concomitantemente.

Para justificar a existência deste módulo fonético, os autores passam a compará-lo com outros módulos existentes na biologia humana³⁷. Um módulo é uma “estrutura neural especial, desenhada para tomar vantagem sistemática, mas apenas de uma única relação entre um *display* proximal de um órgão do sentido e alguma propriedade de um objeto distal” (Lieberman & Mattingly, 1985: p.241). A vantagem de se ter um módulo no sistema é a seguinte:

“Um resultado em todos os casos é que não há, primeiro, uma representação cognitiva do padrão proximal que é modalmente geral, seguido pela tradução de uma propriedade distal particular; ao invés disso, a percepção de uma propriedade distal é imediata, isso significa que o módulo fez todo o trabalho pesado” (Lieberman & Mattingly, 1985: p.241³⁸).

Por exemplo, a localização auditiva é um exemplo de módulo. Uma das várias pistas para a localização auditiva está no intervalo de tempo de chegada, nas duas orelhas, de componentes de uma frequência particular de um dado som. Ninguém pode deixar de argumentar que essa é uma propriedade auditiva geral. Mas os animais tomam vantagem sistemática das relações especiais entre o estímulo proximal e a localização distal. A relação é sistemática por vários motivos. O primeiro se refere ao fato de os animais terem dois ouvidos e, o segundo, de eles estarem localizados separadamente a uma distância. No caso dos humanos, não há tradução de uma disparidade perceptiva porque não há disparidade percebida.

³⁷ Os exemplos dos autores estão presentes em Fodor (1983). Na seção 2.3.3, trataremos dessa obra e será possível verificar a sua influência na teoria.

³⁸ “A result in all cases is that there is not, first, a cognitive representation of the proximal pattern that is modality-general, followed by translation to a particular distal property; rather, perception of the distal property is immediate, which is to say that the moule has done all the hard work” (Lieberman & Mattingly, 1985: p.241).

É diante desse fato que se pode dizer que tanto a localização auditiva dos animais quanto a fala são especiais e cada um tem que ser entendido em seus próprios termos, i.e., em seus respectivos módulos.

A TM também preconiza o fato de o módulo fonético perceptual ser diferente de outros sistemas. Primeiro, com relação aos domínios auditivos e fonéticos há diferença quanto ao local dos eventos distais. Enquanto no domínio auditivo o evento distal está fora e a relação entre o *display* proximal nas duas orelhas é determinada completamente pelos princípios da acústica, o objeto distal no domínio fonético é um gesto fonético (ou o montante de comandos neurais de um gesto do qual um movimento articulatório periférico realiza) e a sua relação com o estímulo proximal terá um traço especial determinado não apenas por princípios acústicos, mas também por um processo muscular interno ao falante. Assim,

“essas intenções são representadas de uma maneira específica no cérebro do falante e há um módulo especializado para conduzir sem esforço o ouvinte a essa representação. De fato, o que é verdade sobre a fala nesse respeito é verdade para todas as línguas, exceto, é claro, que o objeto mais distal da linguagem é alguma representação da estrutura linguística, não meramente um gesto, e tal acesso a esse objeto requer um módulo que não é meramente fonético, mas sim fonológico e sintático” (Liberman & Mattingly, 1985: p.243³⁹).

Segundo, há que se responder como o módulo coopera ou compete com outros que usam estímulos de uma mesma definição física. Basicamente, essa diferenciação ocorre pelas características do estímulo proximal. Por exemplo, para a localização auditiva, o fato de haver diferenças entre a taxa particular de tempo de chegada dos componentes sonoros nas duas orelhas só importa para a localização, i.e. “o *display* proximal e a propriedade distal que os especificam, apenas completam outros aspectos que o ouvinte recebe; eles nunca competem” (Liberman & Mattingly, 1985: p.243). Com relação à percepção fonética, embora as transições da fala compartilhem características de outros sons, nós não “queremos” ouvi-las como glissandos ou trinados, mesmo eles estando em constante “competição direta”.

³⁹ “The particular claim of the motor theory is that these intentions are represented in a specific form in the talker's brain, and that there is a perceiving module specialized to lead the listener effortlessly to that representation. Indeed, what is true of speech in this respect is true for all of language, except, of course, that the more distal object for language is some representation of linguistic structure, not merely of gesture, and that access to this object requires a module that is not merely phonetic, but phonological and syntactic as well” (Liberman & Mattingly, 1985: p.243).

2.2 Como gestos reais

Nesta seção, apresentaremos a seguir uma teoria de percepção da fala que, ao contrário da proposta da TM, toma como primitivos de análise os seus gestos reais. Tomaremos por base a sistematização realizada por Fowler (1996), uma vez que não há um texto que inaugure o modelo tal como há para a TM (Liberman & Mattingly, 1985). No Capítulo 2, desenharemos um quadro mais amplo sobre as motivações iniciais do modelo e seus desdobramentos.

2.2.1 Teoria do Realismo Direto da Percepção da Fala

Fowler (1996), ao escrever sua resposta a Ohala (1996), salienta que a Teoria do Realismo Direto da Percepção da Fala (TRD) discorda quase que em todos os aspectos da TM.

Na sua proposta, primeiramente, os gestos do trato vocal possuem propriedades invariantes próprias e seriam eles mesmos (e não suas estruturas de controles neurais, tal como sugere a TM) os componentes fonológicos de uma sentença: “os gestos fonológicos são ações públicas⁴⁰ do trato vocal que causam estrutura nos sinais acústicos da fala” (Fowler, 1996: p. 1731). Tais propriedades acústicas especificadas pela articulação permitem que a percepção das propriedades fonológicas seja direta, i.e., ser direta significa não ser mediado por processo de teste de hipótese ou inferência e por representações mentais.

Nesta teoria, os ouvintes percebem os gestos porque os sistemas perceptuais têm uma função universal de perceber as causas do mundo real em mídias como a luz, o ar e as superfícies. A autora coloca que a percepção é tida geralmente como heteromórfica com relação à estrutura presente na mídia, i.e., não há relação biunívoca do estímulo distal com relação ao proximal. No entanto, aqui, a percepção não é apenas tomada como homomórfica-*com*, ela é homomórfica-*a-partir*⁴¹, i.e., são os eventos do mundo real que causam estruturas específicas nas mídias. Assim,

⁴⁰ A autora considera os gestos fonológicos como *ações públicas* pelo fato de eles serem reais.

⁴¹ Aqui, a autora faz uma diferença entre *homomorphic with* e *homomorphic of* (Fowler, 1996: p. 1731).

percebemos a fala (e o mundo em geral) porque somos realistas.

Um segundo ponto de discordância com relação à TM reside no fato da TRD não tomar a percepção da fala como especial. Para a autora, não há motivo para se propor um sistema motor de percepção da fala e depois também estabelecer um sistema análogo que realize a percepção visual do sistema de locomoção, uma vez que os dois processos podem ser compreendidos a partir de um mesmo princípio: os seres percebem os movimentos reais, sejam eles referentes à fala ou à visualização dos movimentos envolvidos na locomoção, por exemplo.

De acordo com a proposta, a percepção direta dos gestos reais obedece a funções universais de percepção. Primeiro, para desenvolver sua teoria, a autora argumenta que antes de tudo é necessário embasar os estudos da percepção da fala com o que já se sabe sobre a percepção em geral. Segundo, “a forma que uma teoria de percepção adquire pode ser afetada fundamentalmente pelas assunções que fazemos sobre a produção da fala e a fonologia linguística” (Fowler, 1996: p. 1731), uma vez que os indivíduos que conhecem a fonologia de uma língua são também produtores e percebedores⁴² da fala.

Em outras palavras, o que Fowler vem tentando fazer é diminuir as “distâncias” entre os dados de produção e percepção da fala, de modo que as teorias de produção, de percepção e de fonologia sejam comensuráveis.

Diante disso, a autora tenta, então, enquadrar a percepção da fala em uma teoria de percepção mais ampla. Assim, o que motiva o argumento de que percebemos os gestos reais do trato vocal são as teorias de produção da fala desenvolvidas por Fowler (1993) e Fowler & Saltzman (1993) e de fonologia, que teve as ideias, segundo ela mesma, inspiradas largamente nos trabalhos de Browman & Goldstein (1986, 1992). Assim, como consequência do argumento de que a teoria é direta, real e sobretudo plausível, ela propõe que “embora a percepção dos gestos não seja logicamente necessária para ouvintes e falantes habilidosos, ela é biologicamente necessária” (Fowler, 1996: p. 1732).

De maneira mais detida, Fowler passa a desenvolver em que medida uma teoria mais geral de percepção consegue acomodar os dados da percepção da fala.

⁴² Fowler utiliza a palavra *perceiver*.

Assim, ela consegue justificar o fato dos seus primitivos serem reais e de identificação direta.

Ela parte do fato de uma teoria de aspectos públicos, i.e., que foca o ambiente ou nicho de atuação do percebedor. Tais ambientes provêm informações que dão pistas para a sobrevivência do animal. Por sua vez, as pistas são apreendidas através de investigação empírica e elas são usadas para guiar as ações desses indivíduos.

Essas pistas que informam os percebedores são captadas diretamente pelos sentidos e isso é possível porque os sistemas perceptuais têm funções universais. Primeiro, é através dos sistemas perceptuais que os seres reconhecem seus nichos e “é claro que a evolução os têm moldado para servirem a essa função” (Fowler, 1996: p. 1732). Segundo, os indivíduos usam estruturas em mídia que foram fielmente causadas por eventos no ambiente de modo a veicular a sua fonte de informação. Ela ainda argumenta que os órgãos do sentido não percebem as mídias. “Ao invés disso, essencialmente para sua sobrevivência, eles percebem os componentes do seu nicho que causaram essa estrutura” (Fowler, 1996: p. 1732).

Por fim, a autora embasa sua proposta de que percebemos gestos reais diretamente com o argumento sobre a seleção natural:

“Esses sistemas perceptuais foram moldados pela seleção natural para obedecerem à função de familiarizar os percebedores com os componentes de seus nichos. A percepção auditiva apenas pode ser selecionada para a mesma função. Não há vantagem de sobrevivência em ouvir ar estruturado, mas há vantagem, por exemplo, para localizar um animal pesado fora do campo de visão e para detectar de uma maneira, com relação a si mesmo, que se trata de algo pesado” (Fowler, 1996: p. 1732⁴³).

Como se pode notar nesta breve apresentação da proposta de se adotar primitivos perceptuais como gestos reais, a abordagem de Fowler é de natureza ecológica (influenciada por Gibson 1966⁴⁴), pois tenta dar conta não só das características exclusivas da percepção dos sons da fala. Ela tenta relacionar a sua

⁴³ “These perceptual systems were shaped by natural selection to serve the function of acquainting perceivers with components of their niches. Auditory perception can only have been selected for the same function. There is no survival advantage to hearing structured air, but there is an advantage, for example, to locating a large lumbering animal out of view and to detecting which way, in respect to one's self, it is lumbering” (Fowler, 1996: p. 1732).

⁴⁴ Este trabalho se’ra apresentado na seção 3.2. do Capítulo 2.

teoria a questões de ordem evolutiva, uma vez que (segundo a autora) a perceber diretamente os gestos e o ambiente promove a sobrevivência dos seres.

2.3 Sumário da seção

Nesta seção, apresentamos duas teorias que adotam primitivos de análise de base articulatória. A TM considera como base de sua teoria a articulação pretendida (seus comandos neuromotores) pelos falantes. Por sua vez, a TRD opta por gestos articulatórios reais que são depreendidos diretamente do ambiente com a finalidade de promover a sobrevivências dos indivíduos.

Pretendíamos aqui apenas apresentar as diferentes ontologias de cada teoria para, em seguida (no capítulo 2), abordar como se chegou a esses primitivos a partir dos recortes realizados pelos modelos fonológicos vigentes em cada momento histórico.

CAPÍTULO 2: DEMARCAÇÃO DO TERRENO DAS TEORIAS DE PERCEPÇÃO DA FALA

Em sua “introdução crítica à epistemologia”, Williams (2001) mostra que são cinco os problemas abordados nos estudos epistemológicos. (i) O problema analítico se refere ao conhecimento no que tange, por exemplo, à maneira como diferenciamos o conhecimento da simples crença ou opinião. (ii) O problema da demarcação é subdividido em dois. O problema externo é preocupado com os limites do conhecimento humano, i.e., haveria alguns assuntos passíveis de se conhecer ou apenas teremos alguma opinião (ou fé) acerca deles? Já o interno se ocupa com a demarcação das fronteiras no interior do domínio do conhecimento. Por exemplo, enquanto alguns pensadores defendem a apreensão do conhecimento *a posteriori* ou empiricamente (i.e., dependente da experiência ou observação), outros a tomam independente da experiência, i.e., *a priori*. (iii) O problema do método está relacionado ao modo como obtemos o conhecimento. Ele pode ser abordado com relação às maneiras como as diferentes ciências chegam a seus resultados, ao modo de se “melhorar” a investigação científica e se há métodos de investigação que sejam racionais. (iv) O problema do ceticismo está intimamente ligado ao problema da justificação e ambos estão ocupados com a possibilidade de se poder alcançar o conhecimento. Por fim, (v) o problema do valor se refere de certa forma a todos os outros problemas, uma vez que este só tem significado se houver sentido em possuir conhecimento.

Neste capítulo, trataremos do problema da demarcação, especificamente ao interno. Nas seções seguintes (1, 2 e 3 do Capítulo 2), apresentaremos como os modelos de análise fonológica do Estruturalismo, Gerativismo e de viés dinâmico “loteiam o terreno” da produção e percepção da fala, i.e., apresentaremos quais são os objetos observacionais e teóricos de interesse para cada um deles. No traçar desse histórico, será possível estabelecer uma relação entre a demarcação da área do modelo fonológico com os primitivos de análise das teorias de percepção da fala. A nossa proposta aqui é mostrar que a depender da demarcação do terreno das teorias fonológicas, tem-se diferentes ontologias nas teorias de percepção da fala.

1 A PERCEPÇÃO DA FALA NO ESTRUTURALISMO

Nesta seção, apresentaremos como o Estruturalismo considerava a percepção da fala. Passaremos pela delimitação saussureana do objeto de estudo da linguística moderna e depois pela fonologia estruturalista de base acústica de Jakobson, Fant e Halle (1952) para entender a razão pela qual a percepção da fala pode ser compreendida como audição.

1.1 A Fonologia Estruturalista e o objeto saussureano

Ao delimitar o objeto de estudo da Linguística, Saussure (1969) deixa a fala de fora do estudo da linguagem ao propor a dicotomia *langue/parole*. Nas palavras de Saussure (1969: p.27):

“O estudo da linguagem comporta, portanto, duas partes: uma, essencial, tem por objeto a língua, que é social em uma essência e independente do indivíduo; esse estudo é unicamente psíquico; outra, secundária, tem por objeto a parte individual da linguagem, vale dizer, a fala, inclusive a fonação é psico-física”.

Embora seu corte tenha sido muito mais metodológico do que teórico⁴⁵, vieram com ele implicações epistemológicas interessantes sobre a maneira como se entende a percepção da fala no Estruturalismo. Primeiramente, nesse período fundador, há a proposta de um sistema de conversação entre duas pessoas (A e B), no qual se ilustrava o lugar da língua nos fatos da linguagem (Ilustração 8).

⁴⁵ Não estamos assumindo aqui que Saussure não estava preocupado em fazer uma teoria. Pelo contrário, as suas escolhas fazem parte da sua teoria, i.e., delimitam seu objeto. A sua delimitação é necessária para tornar a linguística uma ciência e sua saída foi a homogeneização do objeto. Nas palavras de Borges Neto (2004: p.52): “Saussure pretende tornar a linguística, verdadeiramente, uma ciência. Para isso, é preciso homogeneizar de qualquer forma o objeto, uma vez que não é possível, no seu entender, descobrir as regularidades necessárias para o estudo científico da linguagem se a linguística não voltar sua atenção para um objeto homogêneo. Apenas a homogeneização do objeto permitirá descobrir nele a sua verdadeira ordem, uma ordem que ultrapasse a mera descrição e que permita chegar ao nível da explicação. A noção de *langue* tem, no quadro da teoria saussureana, o papel de tornar homogêneo o objeto e permitir à teoria linguística aceder à explicatividade.”

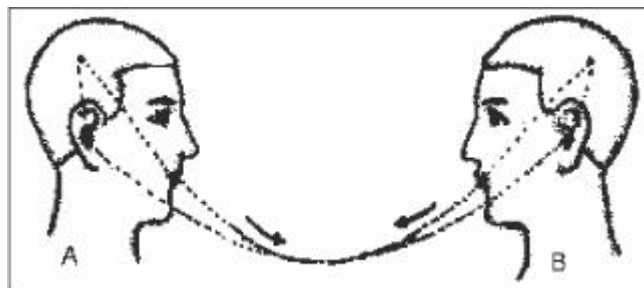


Ilustração 8: Sistema de conversação entre A e B proposto por Saussure (1969 : p.19).

Assim Saussure descreveu o funcionamento de tal sistema (1969: p.19):

“O ponto de partida do circuito se situa no cérebro de uma delas, por exemplo A, onde os fatos de consciência, a que chamaremos conceitos, se acham associados às representações dos signos linguísticos ou imagens acústicas que servem para exprimi-los. Suponhamos que um dado conceito suscite no cérebro uma imagem acústica correspondente: é um fenômeno inteiramente psíquico, seguido, por sua vez, de um processo fisiológico: o cérebro transmite aos órgãos da fonação um impulso correlativo da imagem; depois, as ondas sonoras se propagam da boca de A para o ouvido de B: processo puramente físico. Em seguida, o circuito se prolonga em B numa ordem inversa: do ouvido ao cérebro, transmissão fisiológica da imagem acústica; no cérebro, associação psíquica dessa imagem com o conceito correspondente. Se B, por sua vez, fala, esse novo ato seguirá – de seu cérebro ao de A – exatamente o mesmo curso do primeiro e passará pelas mesmas fases sucessivas”.

Tendo em vista esse sistema e adotando o signo linguístico como unidade de interesse para a Linguística, Saussure acaba por deixar de lado os processos fisiológicos (fonação e audição) e físicos (ondas sonoras). É por isso também que ele coloca que a relação entre a Linguística e a Fisiologia é unilateral “no sentido de que o estudo das línguas pede esclarecimentos à Fisiologia dos sons, mas não lhe oferece nenhum” (Saussure, 1969: p.14). Ainda mais, o funcionamento de seu sistema é garantido por uma “faculdade de associação e de coordenação” entre os conceitos e as imagens acústicas dos signos linguísticos. Assim, na parte psíquica do sistema, ele considera como relações executivas (ativas) o relacionamento de um conceito a uma imagem acústica e, como receptivas (passivas) a associação de uma imagem acústica a um conceito. Em suma, “é ativo tudo o que vai do centro de

associação duma das pessoas ao ouvido da outra, e passivo tudo que vai do ouvido desta ao seu centro de associação” (Saussure, 1969: p.20).

Como se pode notar, toda e qualquer manifestação física e fisiológica referente à produção e à audição da fala é deixada de lado no Estruturalismo saussureano. No entanto, é com os *Preliminaries of Speech Analysis* (Jakobson, Fant & Halle, 1952) que a Fonética passa a ganhar espaço (embora por pouco tempo) dentro da fonologia do período⁴⁶.

Basicamente, a preocupação dos autores era buscar o que havia de minimamente significativo no sinal acústico, isto é, procuravam o que gerava distinção na fala para a transmissão de informação. Assim, os fonemas passariam a ser compostos por feixes de traços de base acústica. Essa medida, em oposição ao sistema saussureano, traria um pouco de Fonética (da parte física secundária) para dentro da Fonologia Estruturalista. Há também uma tentativa de trazer a percepção para a definição dos traços. No entanto, a sua parte no modelo não vai além da descrição sobre como seriam ouvidos os diversos tipos de sons. Um exemplo disso é a “diferença” perceptiva entre vogais e consoantes proposta pelos autores: aquelas teriam força muito maior do que estas.

Diante disso, uma questão nos é colocada: seria a introdução da contraparte física dentro da Fonologia que incentiva um primeiro olhar para a fisiologia na Linguística ou seria a concepção de percepção/audição que faz surgir um interesse pelo funcionamento do sistema auditivo periférico como parte fundamental da percepção da fala no sistema linguístico? Definitivamente, se está em frente a um problema do tipo “ovo/galinha” (Borges Neto, 2004) e esse tipo de problema, normalmente, permanece sem solução. Sua solução (tal como sugere Borges Neto) faz parte do que é importante para endossar sua teoria. Com isso, apresentamos as análises que endossam que o olhar para a fisiologia auditiva advém da fundação da Fonética e Fonologia, pois, como a delimitação da atuação da Fonética e Fonologia

⁴⁶ Estamos considerando a obra de Jakobson, Fant e Halle como parte do Estruturalismo devido ao fato de que, embora os autores tenham assumido o traço como unidade mínima distintiva, há ainda a preservação da noção de fonema, tanto que os traços são responsáveis pela realização das distinções fonêmicas. O fim da noção de fonema somente chega com o *Sound Pattern of English* de Chomsky & Halle (1968). Cabe deixar claro também que a noção de traço não é necessariamente associada ao gerativismo. A fonologia gerativa recebe esse nome por conta da concepção de que uma estrutura fonológica pode ser engendrada a partir de uma estrutura subjacente sobre a qual se aplica um conjunto de regras.

de Trubetzkoy (1939) e a homogeneização do objeto proposta por Saussure garantem cientificidade aos estudos linguísticos, os pesquisadores do período utilizam uma metodologia de análise coerente com essa proposta. Assim, um olhar para a fisiologia da audição vai emergir “naturalmente” ao começar a se questionar como funcionaria a recepção dos sons da fala pelo sistema auditivo periférico.

Passaremos, então, nas próximas seções, às definições de percepção desse período.

1.2 Mas afinal, o que é percepção?

Nesta seção, apresentaremos as definições de percepção da fala disponíveis e vinculadas ao Estruturalismo em Linguística. Primeiro, passaremos por uma justificativa sobre a noção auditiva de percepção. Num segundo momento, para o interesse fisiológico no assunto. Tais pontos, em princípio, endossam a nossa proposta de que o olhar para a fisiologia auditiva advém da dissociação da Fonética e da Fonologia.

1.2.1 Uma metodologia possível

Malmberg (1968), na introdução do seu *Manual of Phonetics*, relata a “pré-história” dos estudos em Fonética. Ele a inicia comentando que a fonética tem seu primeiro desenvolvimento como uma ramificação autônoma na pesquisa linguística em meados do século XIX. Isso, segundo ele, foi consequência da descoberta da *substância* por trás dos símbolos abstratos (que eram as letras). Nas palavras de Malmberg (1968: p.11⁴⁷):

“Um melhor entendimento sobre as substâncias do som foi suposta para dar uma melhor compreensão para as mudanças dos chamados sons da fala. Essa ilusão se tornou o ponto inicial para a fonética instrumental. O novo

⁴⁷ “A better knowledge of the sound substance was supposed to give a better understanding of the changes of so-called speech sounds. This illusion became the starting-point for instrumental phonetics. The new insight into the physiological and physical phenomena which are tied up with human sounds and human sound production, due to the new instrumental methods (specially in physiology), because tools in the search for historical explanations” (Malmberg, 1968: p. 11).

olhar sobre os fenômenos fisiológicos e físicos que estão atados aos sons humanos e a sua produção sonora conduz a novos métodos instrumentais (especialmente em fisiologia), que se tornam ferramentas para a busca por explicações históricas”.

Assim, esse período dito “clássico” da Fonética (tendo como principais nomes Paul Passy, Henry Sweet, Eduard Sievers e Johan Storm) descrevia e analisava os sons das línguas auditivamente. O próprio Malmberg sugere que essas primeiras análises foram colocadas incorretamente como “primariamente unidades fonêmicas conhecidas das línguas dos pesquisadores”, i.e., as análises eram calcadas no próprio conhecimento e impressões que os autores tinham sobre as suas respectivas línguas. Por sua vez, adiciona que esse período contribuiu muito com o conhecimento das variantes dos sons, pois as distinções foram “cuidadosamente estudadas e descritas de maneira acurada”, na maioria das vezes com referência a sua articulação (aberta/fechada, dorsais/palatais, etc). O autor sugere que as análises eram “de fato perceptuais e parcialmente baseadas de maneira inconsciente em interpretações fonêmicas” (Malmberg, 1968: p.12).

No entanto, a principal mudança que ocorre nos estudos em fonética vem do desenvolvimento instrumental para as análises. Tais instrumentos traziam a possibilidade de introduzir os conhecimentos acústicos e fisiológicos sobre os sons e os órgãos da fala. Os resultados com base em análises acústicas não ocorreram em número maior que os com base em dados articulatórios. Tais fatos se devem às possíveis imperfeições de gravação e aos precários dispositivos de análise. Por outro lado, a técnica de base fisiológica atingiu bom desenvolvimento com os laringoscópios, palatogramas e com as figuras de raio-X.

Embora esse desenvolvimento tecnológico tenha contribuído para os avanços sobre a fisiologia e acústica da fala⁴⁸, esse talvez tenha sido também seu maior problema. Malmberg diz que, se por um lado os resultados encontrados foram muito numerosos e foi possível avançar e aperfeiçoar as técnicas de coleta e análise, por outro lado os resultados passaram a ser incompreensíveis de modo que cada vez mais a linguística era invadida por não-linguistas e isso fazia com que os interesses

⁴⁸ Conforme a professora doutora Sandra Madureira (comunicação pessoal), esse período pode ser tratado como uma verdadeira simbiose entre a tecnologia e a ciência da fala. Devido a isso que Malmberg comenta que foram numerosos os resultados encontrados. Contudo, ao que tudo indica, a linguística parece ter se “atrapalhado” com a quantidade de dados coleta.

dos linguistas acabassem sendo deixados de lado. Por exemplo, um pequeno conjunto de parâmetros articulatórios que foi delineado pelos foneticistas do período clássico tinha que ser explicado por uma massa enorme de fatos complicados (tais como lâminas de raio-x e explicações sobre a fisiologia muscular humana) que tinham sido encontrados pelos novos instrumentos. Em suma, “essa enorme massa de variantes descobertas pelos instrumentalistas tinha que ser referida por algum tipo de invariante para ser entendida como elementos em um processo de comunicação” (Malmberg, 1968: p.13).

Diante disso, o conceito de fonema foi formulado e frequentemente tomado como uma unidade estável; uma classe ou uma abstração das quais as manifestações físicas poderiam ser interpretadas como variantes.

Malmberg coloca que, infelizmente, os foneticistas do período clássico propuseram (com base na glossemática de Hjelmslev⁴⁹) a divisão de que a Fonética seria uma ramificação das ciências naturais, uma ciência da substância, enquanto o estudo das funções do fonema (a Fonologia) seria uma ciência humanística. Vale acrescentar que não foram só os foneticistas que definiram a Fonética e a Fonologia dessa maneira. Trubetzkoy (1970[1939]) dá a cada uma das duas disciplinas essa mesma definição nas páginas iniciais do seu *Princípios de Fonologia*.

É interessante notar que o autor tem um posicionamento insatisfeito com a divisão entre Fonética e Fonologia, pois ele sugere que “a expressão linguística – apenas a expressão com conteúdo linguístico – é forma e substância” (Malmberg, 1968: p.14). Para ele a divisão foi muito mais metodológica⁵⁰, pois a individualização das duas disciplinas foi muito mais um retorno para os aspectos estritamente linguísticos, isto é, decidiu-se concentrar os olhares sobre as características relevantes dos sons da língua. Tendo isso em vista, ele coloca que as oposições sugeridas pelos linguistas do Círculo Linguístico de Praga são foneticamente baseadas e esses traços relevantes são traços de substância.

Por fim, o autor sugere que a percepção tem que ser audição. O linguista teria que utilizar a audição para tomar a substância de um dado som, pois haveriam que

⁴⁹ Uma boa introdução sobre a glossemática de Hjelmslev se encontra em ILARI, R. (2005). O estruturalismo linguístico: alguns caminhos. In: MUSSALIN & BENTES (Org). *Introdução à linguística: fundamentos epistemológicos*. Volume 3. 2ª edição. São Paulo: Cortez, 2005.

⁵⁰ Em realidade, trata-se de uma divisão epistemologicamente motivada.

ser consideradas apenas as características distintivas de um sistema. Assim, pode-se dizer que, coerentemente, os pesquisadores buscavam mostrar as diferenças que eles ouviam com a ajuda de instrumentos.

1.2.2 A fisiologia da audição como linha promissora de pesquisa

Como se viu na seção anterior (1.1), devido a uma decisão epistemológica, os estudos em fonética têm a audição como “instrumento” de análise. Assim, desenvolvem-se nesse período trabalhos com o interesse no funcionamento fisiológico do sistema auditivo periférico e interno.

Basicamente, Brosnahan & Malmberg (1970: p.160) relatam o estado da arte nos estudos de percepção do período Estruturalista. Eles a definem “como um evento da fala que toma lugar no ouvinte, a percepção dos distúrbios de uma onda sonora na atmosfera, a identificação desses distúrbios e as interpretações dela”. Vale notar que a maior preocupação do capítulo é mostrar minuciosamente o funcionamento dos ossos e membranas envolvidas no processo de audição. Aqui, o mais interessante é notar que embora os autores conhecessem a fundo o funcionamento auricular humano, eles tinham em vista que os resultados dos trabalhos experimentais da percepção da fala ainda não eram completamente entendidos. Grosso modo, Brosnahan & Malmberg não sabiam como o cérebro interpretava e identificava os sons da fala como um todo. E o princípio de associação proposto por Saussure aparentemente não garantia a percepção dos sons da fala (como veremos na seção 2 do Capítulo 2).

Após apresentar o processo de audição⁵¹ (que é basicamente preocupado com os limites da audição humana), os autores passam aos resultados experimentais. Tais resultados não passam de especulações que sugerem que a percepção auditiva da fala é uma linha de pesquisa muito promissora. Também é interessante notar que eles a definem como misteriosa. Assim, esses aspectos levantados anteriormente sugerem a falta de uma teoria sobre a percepção da fala.

⁵¹ Brosnahan & Malmberg apresentam o funcionamento do ouvido humano, mas a principal preocupação nos estudos resenhados por eles é descobrir quais são os limites da audição, i.e., qual é a taxa de frequência que as pessoas ouvem e onde se encontra o limiar da dor, por exemplo.

Por fim, após sugerir trabalhos interessantes a serem desenvolvidos (como explorar os efeitos de interferência no que ouve o receptor com uma onda de fala não analisada sobre o sinal de fala e pensar como são ouvidos os seus componentes e traços distintivos), conclui-se que a fala é muito resistente à distorção. Para os autores, devem-se estudar as propriedades auditivas que garantem sua inteligibilidade. Eles ainda sugerem que ela é garantida pela redundância presente no sinal acústico.

Tal como foi mostrada nas seções anteriores, a tentativa de entender como funciona a audição e de traçar uma agenda de estudos em percepção está de acordo com a concepção de fonética do Estruturalismo, i.e, os pesquisadores a consideram como parte física e fisiológica no processo de comunicação, de modo que eles apenas tentam desvendar experimentalmente os fenômenos identificados como distintivos.

1.3 Sumário da seção

Nesta seção apresentamos como a delimitação de um novo objeto para os estudos linguísticos a partir dos trabalhos de Saussure determinou o desenvolvimento de uma fonologia preocupada apenas com os aspectos distintivos da fala. Tais abordagens endossaram uma perspectiva auditiva nos estudos de percepção, uma vez que as características distintivas do sistema fonológico de uma dada língua deveriam estar presentes no sinal acústico.

Tal como mostraremos a seguir, essa abordagem auditiva foi sendo substituída à medida que novos dados sobre a produção da fala passaram a ser contemplados dentro de novas teorias fonológicas, como o Gerativismo (seção 2, a seguir), que, por sua vez, embasaram uma proposta de teoria de percepção da fala.

2 A PERCEPÇÃO DA FALA NO GERATIVISMO⁵²

Nesta seção, apresentaremos o percurso da Teoria Motora da Percepção da Fala (doravante TM). Partiremos da sua primeira formulação e, em seguida, passaremos por uma “correção no meio do caminho”. Por fim, trataremos da sua versão revisada de 1985.

Esse percurso mostrará que a teoria ganha maior força (ou poder explicativo) quando incorpora as premissas da Fonologia Gerativa de Chomsky & Halle (1968), os achados referentes ao Efeito McGurk (McDonald & McGurk, 1976) e da Modularidade da Mente de Fodor (1983).

2.1 Uma tentativa de colocar tudo junto: a “*Early Motor Theory*”

Liberman (1957) faz um relato dos resultados alcançados nas pesquisas em percepção de fala realizadas nos Laboratórios *Haskins* desde o início da década de 1950. É interessante notar que os dados que embasaram essa primeira proposta de uma teoria de percepção da fala com base articulatória foram encontrados por acaso, uma vez que esse grupo de pesquisadores estava utilizando uma metodologia experimental e os dados os levaram a interpretações inesperadas mesmo utilizando uma “concepção” auditiva de percepção.

Primeiro, caracterizava-se por um experimentalismo devido à utilização de uma máquina (o *Pattern Playback*⁵³) que sintetizava os sons através da leitura óptica de espectrogramas desenhados à mão. Essa maneira prática de sintetizar a fala tinha dado fácil acesso a várias partes significativas do padrão acústico do sinal de fala, pois era possível isolar e testar padrões que eram tidos como pistas acústicas para a identificação de certos sons da fala. O próprio Liberman afirmava que muitos dos avanços que se obtiveram com o *Pattern Playback* ocorreram na tentativa e

⁵² A proposta de que a Teoria Motora de Percepção da Fala de Liberman & Mattingly (1985) está visivelmente vinculada a um arcabouço Gerativista só ocorre a partir da proposta da Fonologia Gerativa (Chomsky & Halle, 1968) e da Modularidade da Mente (Fodor, 1983). Antes desses dois marcos teóricos, as propostas da Teoria Motora estão vinculadas ao arcabouço estruturalista. Inserimos essas versões iniciais do modelo nesta seção com o intuito de traçar o percurso da teoria que sai do Estruturalismo e chega ao Gerativismo.

⁵³ Para maiores informações sobre o funcionamento do *Pattern Playback* ver a Introdução e Capítulo 2 de Liberman (1996).

erro.

Segundo, a tentativa de sintetizar os sons da fala com o auxílio de espectrogramas partia da análise auditiva dos resultados produzidos. Isto é, os juízos de aceitabilidade e de identificação das produções sintéticas eram realizados pelos próprios pesquisadores e, quando essas produções eram utilizadas em experimentos com outros informantes, os resultados mostravam, por exemplo, que a identificação das categorias era mais estável dentro das fronteiras fonêmicas da língua dos informantes. Isso corroborava a proposta de se buscar a “substância” apenas nas produções distintivas, ou seja, nas produções que apresentavam relevância linguística (tal como apontamos nas seções anteriores). No entanto, essa busca por pistas perceptuais dentro dos padrões fonêmicos é que acabou gerando problemas para a abordagem auditiva de percepção da fala.

Basicamente, os estudos com o *Pattern Playback* tinham conseguido realizar generalizações interessantes sobre o comportamento dos componentes acústicos da fala. Por exemplo, havia-se conseguido mostrar que as transições do primeiro formante indicavam pistas para o modo de articulação dos sons. Por sua vez, as transições do segundo formante eram pistas para o ponto de articulação.

Em 1957, Liberman consegue uma generalização ainda mais interessante: as transições do segundo formante apresentavam posições de frequência características (ou *loci*) nas quais começariam ou estariam apontadas.

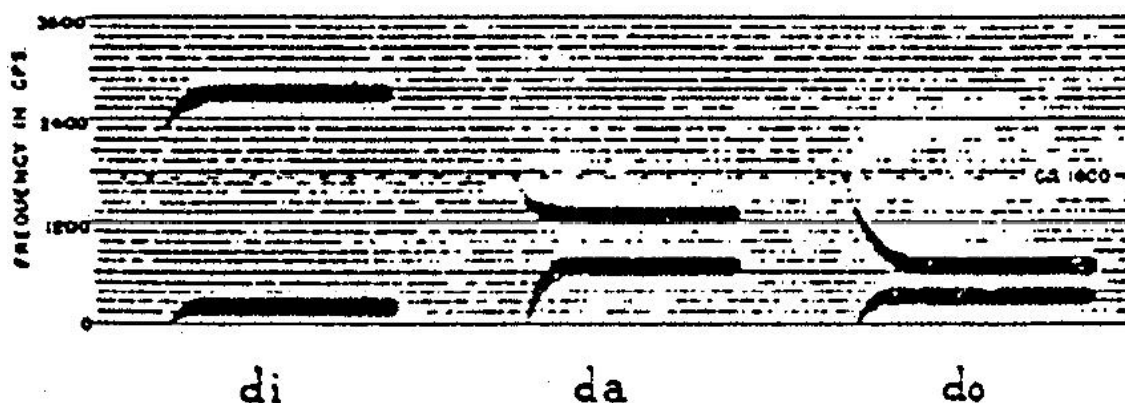


Ilustração 9: Padrões espectrográficos que produzem /di/, /da/ e /do/. As linhas pontilhadas mostram o locus de 1800cps para o /d/ (Liberman, 1957 apud Liberman, 1996: p.190).

Na Ilustração 9, nota-se que a transição do segundo formante sempre vem da (ou aponta para) região de 1800Hz. Os autores notaram isso, pois nos primeiros experimentos o “mistério” pairava sobre a grande variabilidade da transição do segundo formante com relação à qualidade da vogal que seguia a consoante. Isto porque naquele momento não se conseguia encontrar nenhuma relação entre o acústico e o articulatório, uma vez que por mais que sempre se identificasse a consoante de início de sílaba como /d/, por exemplo, o sinal acústico nem sempre era igual, pois havia uma variação na transição a depender da vogal de núcleo silábico. Essa generalização sobre a existência de um *locus* para cada consoante dava conta da aparente falta de correspondência de um para um entre a informação acústica veiculada no ar e a ação articulatória dos órgãos do trato vocal.

Entretanto, essa generalização apenas era possível para o /d/. Ao buscar o mesmo fenômeno com outras consoantes, não houve um *locus* regular para /g/ antes de todas as vogais, por exemplo. Na ilustração 10, temos as transições da consoante /d/ antes de outras vogais. Pode-se notar que as transições parecem apontar para 1800Hz em todos os ambientes vocálicos.

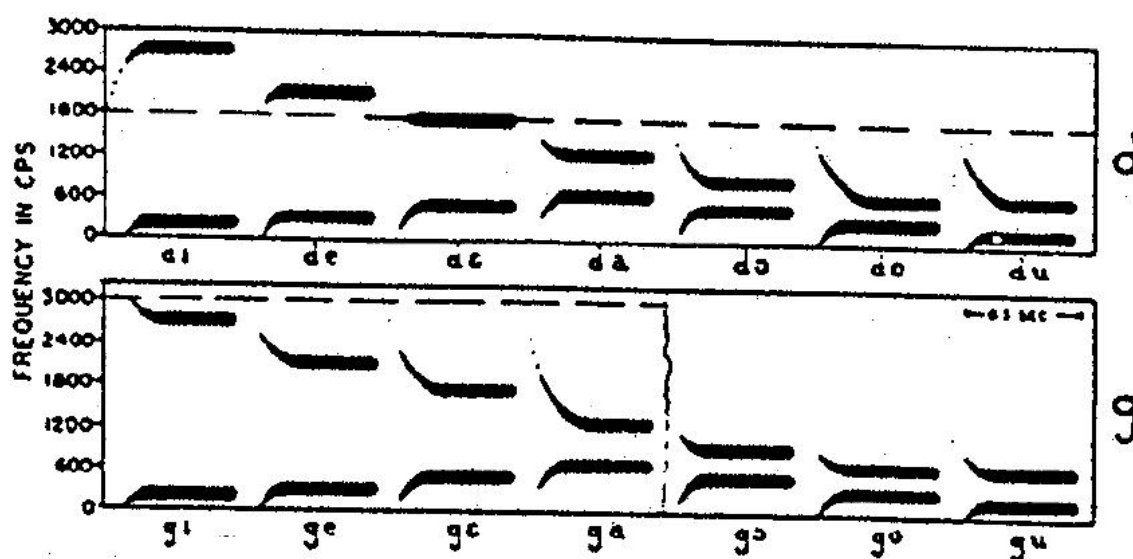


Ilustração 10: Padrões espectrográficos que produzem /d/ e /g/ em vários contextos vocálicos. As linhas pontilhadas indicam os loci das consoantes (Liberman, 1957 apud Liberman, 1996: p.194).

A situação, por sua vez, não é igual para /g/. Nota-se que as transições parecem partir de 3000Hz e vão se tornando mais longas até /ga/. No entanto, a partir de /gɔ/ a transição parece perder o seu *locus*, pois fica menos acentuada e mais breve. Liberman (Liberman, 1957 *apud* 1996: p.194) coloca que nesses dados

“nós temos uma descontinuidade real no nível acústico. Nós não somos capazes de encontrar invariantes acústicos para corresponder [...] à falta de troca na percepção de /g/ nesta série, e há níveis racionais para supor que não exista algum. O importante, claro, é que esta descontinuidade no nível acústico não é emparelhada por alguma descontinuidade correspondente na articulação ou na percepção. A articulação do /g/ é essencialmente a mesma em todos os exemplos, e assim também a percepção”⁵⁴.

Tendo esses resultados em vista, o autor faz a seguinte pergunta: “quando articulação e onda sonora vão por caminhos separados, por qual a percepção iria?” (Liberman, 1957 *apud* Liberman, 1996: p.193-194) Para ele, a resposta é muito clara: a percepção sempre segue a articulação. O autor sugere isso devido às transições de /g/, pois se elas identificassem sistematicamente um *locus* acústico para determinadas vogais (ver ilustração 10) seria evidente que a percepção seria guiada pelo acústico, uma vez que são as próprias ondas acústicas que chegam aos ouvidos. Ainda mais, se a percepção dependesse mais diretamente dos estímulos acústicos, então as pessoas teriam que realizar mudanças articulatórias para manter um padrão sonoro constante. No entanto, isso não parece ocorrer.

Entretanto, como ficariam os estudos em fonética frente a esses dados, sendo que a concepção de percepção presente nesse período era auditiva? Como as pessoas “cortariam” o sinal acústico em “pedaços articulatórios” para aprender uma língua?

Liberman sugere que o processo de percepção não poderia ser levado em conta apenas sobre as características acústicas da fala. Teria que haver a possibilidade de se considerar estímulos de ordem proprioceptiva. A saída, então,

⁵⁴ “[...], we have a real discontinuity at the acoustic level. We have been able to find no acoustic invariant to correspond to – or, if you will, to explain – the unchanging perception of /g/ in this series, and there are reasonable grounds for supposing that none exists. The important thing, of course, is that this discontinuity at the acoustic level is not paralleled by any corresponding discontinuity in articulation or in perception. The /g/ articulation is essentially the same throughout, and so also is the perception” (Liberman, 1957: p. 194).

para explicar como os movimentos articulatórios e seus efeitos mediados por estímulos acústicos seriam percebidos é através de uma “extrema e antiquada forma”: “esse olhar diz que nós notoriamente imitamos [*mimic*] os sons da fala enviados e então respondemos com estímulos tácteis e proprioceptivos que são produzidos pelos nossos próprios movimentos articulatórios” (Liberman, 1957: p.195).

Em princípio, pode-se dizer que sim, se se levar em conta que o behaviorismo ainda era a teoria de aprendizagem corrente nessa época, pois *Review Of Verbal Behaviour by B.F. Skinner*, feita por Chomsky, apenas seria publicada em 1959. No entanto, dir-se-ia não, se se considerar que uma das propostas para os traços distintivos era de base acústica (Jakobson, Fant & Halle, 1952). Haveria, então, que se reconsiderar a natureza dos traços para que fosse feita uma ponte entre a produção e a percepção na fonologia das línguas.

Ainda mais, se se levar a discussão além, pode-se notar que em realidade o Estruturalismo não dá conta dos dados. Seria possível, no Estruturalismo, desenvolver uma teoria de percepção da fala sendo que a concepção (tanto teórica quanto metodológica) de percepção é auditiva e os dados experimentais encontrados mostram justamente o contrário?

Quando se leva em consideração o viés behaviorista para dar conta dos dados, na verdade ele apenas consegue dar conta da categorização dos sons. A proposta de que se imita e assim se consegue chegar aos mesmos padrões acústicos não explica a percepção do som em si. Apenas é explicada a sua categorização.

Vale lembrar também que os dados apresentados por Liberman (1957) são uma primeira tentativa de dar uma interpretação teórica aos achados dos Laboratórios *Haskins*. Essa tentativa é considerada como a primeira versão da Teoria Motora da Percepção da Fala (a chamada *Early Motor Theory*). No entanto, como já foi mostrado nas linhas anteriores, os dados apenas apontam para uma preferência da articulação sobre o acústico e, com isso, se propõe uma saída para a categorização dos sons.

Em suma, nessa primeira tentativa de acomodar os achados das pesquisas

realizadas nos Laboratórios *Haskins* se encontra a proposta de que os primitivos da percepção da fala são de base articulatória, devido à falta de invariância do sinal acústico; havendo, com isso, uma falta de correlação entre a articulação dos órgãos do trato vocal e as suas características acústicas.

2.2 Uma correção no meio do caminho

Lieberman e colegas (1967) publicam um texto que foi, nas próprias palavras de Liberman (1996: p.201), “o mais frequentemente citado, mas infelizmente pelo efeito e motivo errados”, pois nele se encontram os dados referentes ao que se chamou de percepção categórica. Antes de se passar aos dados discutidos no trabalho, é importante mencionar que nesse período a teoria passa por reformulações que mudam a perspectiva (ou direção) do olhar sobre a percepção da fala. Em outras palavras, abandona-se o olhar horizontal presente na *Early Motor Theory* para se assumir uma orientação vertical. Enquanto na primeira formulação da teoria a percepção é tida como “um processo auditivo geral, seguido pela indexação às consequências sensoriais de gestos⁵⁵ de acordo com um processo igualmente geral de aprendizagem por associação” (i.e., não havia nada biologicamente ou psicologicamente distinto na percepção da fala com outros sons), na formulação de 1967, não se assume dois estágios para a percepção. Ela é tida como “um processo, com ligações apropriadas entre componentes motores e sensoriais” e como “processos de interdependência perceptual e produtiva”. Ainda mais, postula-se que essa interdependência não ocorre como uma instância de algum processo geral de ação e percepção, ele é sobretudo um sistema distinto chamado de “decodificador de fala” (Lieberman, 1967: p.201) ou “módulo da fala”. O caminho percorrido até chegar a essa proposta não é tão curto quanto parece. Passaremos aos argumentos que a sustentam no decorrer desta seção.

Diante do fato de a fala apresentar uma “marcada falta de correspondência entre o som e um fonema percebido” (Lieberman e colegas, 1967: p.203), a pergunta central do texto de 1967 é: por qual mecanismo o ouvinte decodifica os sons e

⁵⁵ A utilização de termo *gesto articulatório* aqui apenas se relaciona aos movimentos dos articuladores no trato vocal. Não há correspondência com a proposta de um novo primitivo de análise presente em Browman & Goldstein (1986 e seguintes).

recupera os fonemas? Para responder a essa questão, os autores investigaram quatro pontos que embasam a proposta de se assumir a existência de um módulo de fala. Trato deles a seguir.

2.2.1 Pode ser a fala alfabética?

Os autores defendem a hipótese de que os fonemas não podem ser tomados como um alfabeto sonoro. Essa questão é importante para que desconstrua a noção de que a organização fonológica das línguas pode ser algo parecido com a organização alfabética da escrita⁵⁶.

O principal argumento reside na própria falta de correspondência entre sons e fonemas, que impossibilitaria a sua recuperação através dos sons. Ainda mais, um alfabeto sonoro não atingiria a efetividade comunicativa que se encontra na fala, uma vez que há um limite para o ouvido humano na “quantidade” de sons que podemos codificar. Segundo Orr, Friedman & Williams (1965 *apud* Liberman e colegas, 1967: p.204), a fala é codificada com dificuldade a partir de 400 palavras por minuto e, se consideramos que em média há 4 ou 5 fonemas por palavra da língua inglesa, temos uma taxa de 30 fonemas por segundo.

Esta mesma taxa de sons discretos foi apontada por Miller & Taylor (1948 *apud* Liberman e colegas, 1967: p.204) como desencadeador de um ruído impossível de se analisar. O ouvinte até consegue captar o *pitch* da sentença, mas o seu significado não. Esses estudos ainda apontam para o fato de que até mesmo 15 fonemas por segundo (o que já é usual numa conversa) é mais do que o ouvido suporta se os fonemas forem considerados como uma cadeia de eventos acústicos discretos.

Temos que considerar que outros alfabetos sonoros utilizados não chegam perto da efetividade da fala. O código Morse internacional, por exemplo, é um tipo de criptograma sonoro que funciona de maneira pobre em comparação com a fala, até

⁵⁶ O motivo pelo qual essa questão está nesse trabalho de 1967 não está muito clara. Provavelmente, trata-se de alguma discussão na qual os autores do período se faziam. O que é interessante notar é que um ano após esse texto o conceito de fonema é abandonado a partir da proposta de Chomsky & Halle (1968) no *Sound Pattern of English*. Não estamos argumentando que os autores se valeram das discussões de Liberman e colegas (1967) e sim de que, com alguma certeza, essa era uma questão que estavam na ordem do dia.

mesmo nos casos em que seus usuários têm anos de prática. Ainda menos eficiente (nesse período), são as máquinas desenvolvidas para converter textos em fala para os cegos. A maior dificuldade dessas máquinas não é converter o material impresso em textos orais, e sim as limitações perceptuais dos próprios usuários.

Por fim, não há vantagem em se considerar os fonemas como alfabetos sonoros, pois há diferenças significativas entre ler e escrever. Os autores argumentam que ler é uma tarefa simples, pois se trata de um código simples. No entanto, ouvir se relaciona a uma tarefa complexa que, mesmo assim, é realizada sem dificuldades pelos ouvintes. Também ressaltam que enquanto ouvir é uma tarefa universal e básica para as línguas, ler um alfabeto não é.

2.2.2 A reestruturação dos fonemas no som

Para dizer que a fala é um código de fonemas reestruturado no sinal acústico é preciso descobrir quais são as características que subjazem à sua contraparte sonora e que mostram que a fala é um código complexo. Neste ponto, os autores vão buscar quais são as pistas acústicas que compõem o fonema durante a produção do som e vão mostrar, a partir de dados do segundo formante de /d/, como a reestruturação ocorre do fonema para o som. Quatro características serão apresentadas.

A primeira característica que mostra a complexidade do sinal acústico é a variação das pistas acústicas a depender do contexto em que se inserem. Na ilustração 11, a seguir, é possível notar que as transições do segundo formante são diferentes para /di/ e /du/ e que mesmo assim se percebe um /d/ nas duas sílabas. Os próprios autores argumentam que ao se isolar apenas as transições, há a produção de sons distintos: um assobio ou glissando ascendente para /di/; e um descendente para /du/.

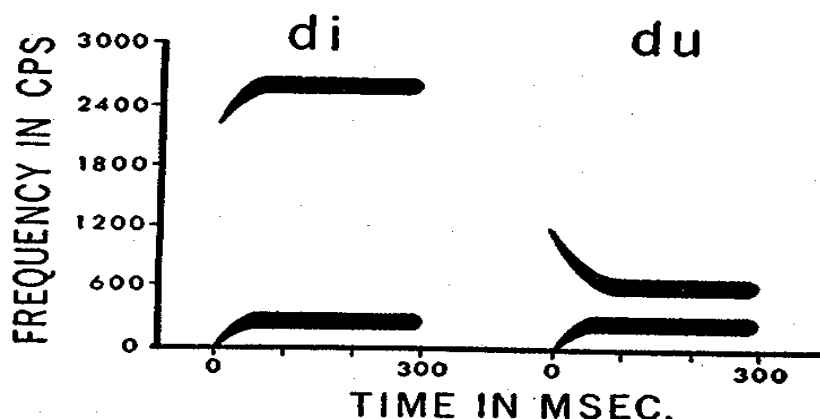


Ilustração 11: Padrões espectrográficos suficientes para sintetizar /d/ diante de /i/ e /u/ (Liberman e colegas, 1967: p.206).

Uma segunda característica apontada pelos autores está no fato de que na cadeia da fala há o desaparecimento da fronteira fonêmica, de modo que o “sinal da fala, tipicamente, não contém segmentos que correspondem ao discreto e à comutação de fonemas” (Liberman e colegas, 1967: p.207), i.e., não se corta um segmento e o substitui (como ocorre na escrita). Surge, então, a proposta de que a transição da consoante para a vogal dá informação de dois fonemas ao mesmo tempo, ocorrendo uma transmissão em paralelo. Por exemplo, se cortarmos uma sílaba da direita para a esquerda, apenas ouviremos ou a sílaba ou um ruído (quando estiver acabando a vogal). Nunca se ouvirá apenas a consoante. Assim, a transição transmite a informação dos dois fonemas.

O terceiro ponto levantado pelos autores diz respeito ao *locus* das consoantes. Delattre, Liberman e Cooper (1955) vão propor que há um *locus* para cada consoante, mas a questão é se ele é um invariante acústico. Para os autores, o *locus* é a região para qual as transições de uma consoante convergem. Se tomarmos o exemplo da ilustração 11, vemos que /di/ e /du/ apresentam transições ascendente e descendente, respectivamente. Mas se considerarmos outros contextos (como em 10), veremos que as transições convergem para um mesmo ponto, que para /d/ é de 1800Hz: trata-se do *locus* de /d/.

No entanto, se considerarmos a Ilustração 12A, a seguir, notamos que, se as transições saírem todas da mesma frequência, temos o problema da falta de

correspondência entre o sinal acústico e o fonema, de modo que temos a produção de /b/, /d/ e /g/. Em contrapartida, em 12B temos a inserção de um intervalo de silêncio que promove a produção de /d/ apenas. Os autores, então, argumentam que a noção de *locus* é uma noção muito mais articulatória do que acústica, pois ele não está presente no sinal acústico. Stevens & House (1956 *apud* Liberman e colegas, 1967: p.209) apresentam cálculos que mostram que no momento da oclusão a frequência é de 1800Hz, mas nenhum som emerge até que haja alguma abertura: “a frequência do *locus* não é radiada como parte do sinal acústico”.

Tais fatos endossam a proposta dos autores de que a invariância está na articulação e não no som da fala: “melhores aproximações para a invariância provavelmente são anteriores na cadeia dos eventos articulatórios, indo além do formato da cavidade que subjaz ao *locus* para os comandos que produzem os tais formatos” (Liberman e colegas, 1967: p.209). Como veremos mais adiante, a invariância está em realidade nos comandos musculares.

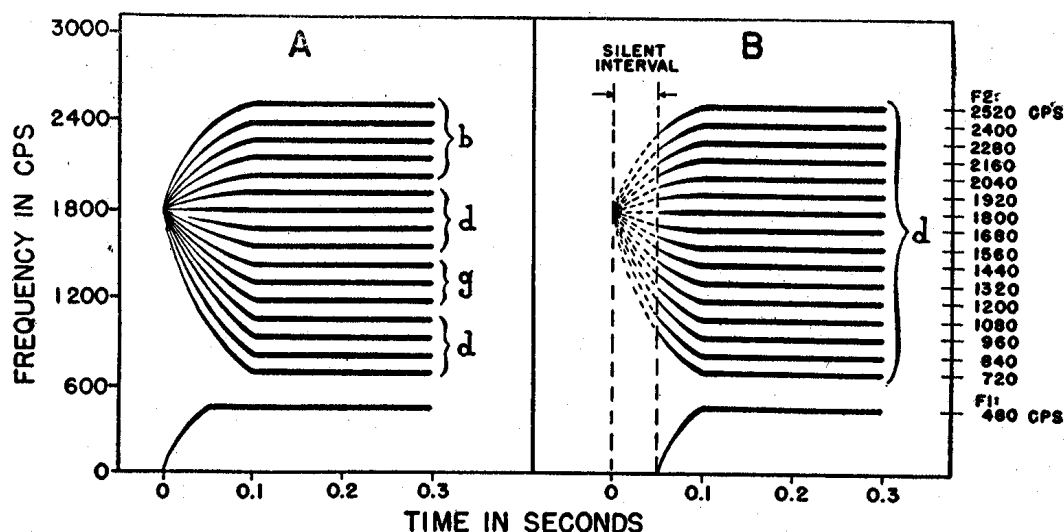


Ilustração 12: (A) Transições que se iniciam no locus de /d/. (B) Transições que apenas apontam para o locus de /d/ (Liberman e colegas, 1967: p.208).

Na tentativa de sistematizar a proposta de que “os fonemas segmentais são drasticamente reestruturados no nível do som” (Liberman e colegas, 1967: p.209), os autores vão comentar que, quando ampliamos a análise sobre o segundo

formante de outras consoantes (como /k/ e /g/), a situação fica ainda mais “complicada”, pois a falta de correspondência entre o sinal acústico e o fonema percebido é muito maior que o encontrado para /d/. Por exemplo, uma mesma consoante tem como pista acústica transições distintas a depender do contexto em que se inserem. Por exemplo, uma transição de 1440Hz é ouvida como /p/ antes de /i/, mas como /k/ antes de /a/.

Outros parâmetros como modo, vozeamento e posição na sílaba também apresentam problemas, pois “as pistas de modo, ponto e vozeamento de /b/ em /ba/ são acusticamente diferentes do /b/ em /ab/, as pistas transicionais são quase imagens espelhadas; ainda mais, o conjunto precedente difere dos conjuntos correspondentes para /b/ em cada um dos contextos com outras vogais” (Liberman e colegas, 1967: p.211).

Com relação às vogais temos comportamento totalmente oposto às consoantes, i.e., a sua percepção parece ser guiada pelos formantes, pois não há variação a depender do contexto em que se inserem. No entanto, os experimentos com fala sintetizada “criam” um contexto de produção que não ocorre na fala normal, já que no seu uso praticamente não há um estado estacionário tão longo quanto o produzido experimentalmente. Os autores sugerem que essa “invariância acústica” se deva a isso.

Por fim, os autores tratam do problema da segmentação da cadeia sonora para endossar a sua proposta de que os fonemas não são encontrados no nível acústico, pois

“se alguém examina as pistas acústicas mais geralmente, ele encontra que sucessivos fonemas são mais comumente localizados fundidos no sinal acústico. Isto é, como veremos, trata-se de um correlato do processamento em paralelo que caracteriza o código da fala e é uma condição essencial para a sua eficiência. Uma consequência é que as pistas acústicas não podem ser divididas em segmentos do tamanho de um fonema no eixo do tempo (Liberman e colegas, 1967: p.211)⁵⁷.

Esse aspecto é endossado pelo trabalho de Harris (1953 *apud* Liberman e

⁵⁷ “If one examines the acoustic cues more generally, he finds that successive phonemes are most commonly merged in the sound stream. This is, as we will see, a correlate of the parallel processing that characterizes the speech code and is an essential condition of its efficiency. One consequence is that the acoustic cues cannot be divided on the time axis into segments of phonemic size” (Liberman e colegas, 1967: 211).

colegas, 1967: p.211), no qual ele argumenta que a montagem de palavras a partir de segmentações de palavras já existentes como se fossem “tijolos” (i.e., realizou-se o recorte e a colagem de consoantes e vogais previamente gravadas) produz palavras que soam inteligíveis e sem naturalidade. Para eles, esse é motivo pelo qual não tem sido possível sintetizar fala a partir de segmentos pré-gravados (naquele período). À guisa de conclusão, os autores também apresentam o fato de que a “segmentação também exhibe, de um lado, uma relação completa entre a estrutura linguística e percepção e, de outro lado, com a cadeia sonora” (Liberman e colegas, 1967: p.212), ou seja, a cadeia sonora deve ser mediada por uma estrutura linguística (o fonema) para conseguir ser percebida e produzida.

2.2.3 O módulo da fala

Na seção anterior (2.2.2), apresentamos alguns dados que apontam para a proposta de que os fonemas são reestruturados no sinal acústico. Diante do fato de que há perda da correspondência entre o sinal acústico e o fonema e de que mesmo assim os fonemas são produzidos de maneira silabada na cadeia da fala, os autores vão sugerir que deve haver um decodificador apropriado para a fala. Para sustentar essa proposta, eles vão expor experimentos que mostram que a percepção da fala é diferente da de outros sons⁵⁸.

Segundo o próprio Liberman (1996: p.201), os dados sobre o que se convencionou chamar de percepção categórica é de longe o fato mais citado sobre a percepção da fala. Para ele, os dados sobre a percepção categórica não são os mais importantes para a sustentação da Teoria Motora, no entanto os resultados apontam para o fato de se considerar a existência de um módulo da fala.

Nesse experimento, os informantes foram submetidos a variações contínuas de consoantes (variação do segundo formante) e de vogais, de modo que eles tinham que assinalar quando havia a mudança do som a uma nova categoria. Os resultados encontrados são distintos para as consoantes e para as vogais. Para aquelas, os ouvintes não notaram variações “aos poucos”, i.e., eles ouviam “saltos

⁵⁸ Para os autores, a percepção da fala é especial a ponto de se exigir estratégias cognitivas distintas das envolvidas na percepção de outros sons. Por exemplo, ouvir a fala é diferente de ouvir uma buzina de carro, uma serra elétrica ou uma porta batendo.

quânticos” de uma categoria a outra. Já para estas, os ouvintes conseguiam identificar que as vogais variavam passo a passo.

Esses dados mostraram para os autores que conseguimos discriminar muitos mais sons do que podemos identificar, ou seja, há a percepção dos fonemas reestruturados, uma vez que se ouvem os fonemas mas não as suas variações intrafonêmicas. Nas palavras deles: “a percepção imediata é mais próxima do contínuo do que do categórico e a pessoa ouve muito mais variações intrafonêmicas [...], ouvintes podem discriminar muito mais estímulos do que podem⁵⁹ absolutamente identificar (Liberman e colegas, 1967: p.213).

Mas esses dados endossam então que a percepção é ou não é categórica? A explicação apresentada pelos autores é que a diferença na percepção de vogais e consoantes pode ter ocorrido devido ao desenho do experimento, pois, como os estímulos das vogais foram apresentados isoladamente, os informantes podem ter interpretado esses sons como não sendo fala. Esse ponto endossa a proposta da existência de um módulo para a fala, pois se interpretamos um som qualquer fora desse módulo, ele não será interpretado como fala e, assim, estará sujeito a mais discriminações. Conforme Stevens (1966 *apud* Liberman e colegas, 1967: p.213) se as transições dos sons forem rápidas, a tendência é de que a percepção seja categórica. Isso explicaria a diferença de resultado entre consoantes e vogais, de modo que quanto mais dinâmico for um som, mais identificado como fala será. Em suma, esses resultados refletiam a operação do módulo da fala.

Por fim, os autores ainda apresentam mais um fato que endossa a proposta da existência de um módulo da fala. Eles argumentam que o processamento da fala ocorre em hemisfério do cérebro diferente do que para outros sons: “o sinal da fala codificado é mais rapidamente decodificado no hemisfério esquerdo do que no direito” (Liberman e colegas, 1967: p.214).

2.2.4 Codificação e decodificação

Tendo então argumentado a favor do fato de a fala ser a versão acústica de

⁵⁹ Aqui caberia questionar os autores com relação ao fato de os seres humanos poderem ou precisarem identificar mais categorias de que realmente fazem.

fonemas reestruturados e de que é preciso um decodificador para percebê-la, os autores vão explicitar como ocorre o processo de codificação (movimento do fonema para o sinal acústico) e de decodificação (do sinal acústico para o fonema). Cabe ressaltar que essa proposta tenta dar conta dos dois processos fazendo uso de uma “via de mão dupla”, de modo que tanto produção quanto percepção sejam mediados por um único processo que vai do fonema para o som e do som para o fonema. Esta ressalva se faz necessária para que o modelo seja mais elegante uma vez que

“não parece parcimonioso assumir que o ouvinte-falante aplica dois processos inteiramente diferentes às situações de *status* igual, um para codificar a linguagem e outro para decodificá-la. Uma assunção mais simples é que há apenas um processo, com ligações apropriadas entre os componentes sensoriais e motores” (Liberman e colegas, 1967: p.221)⁶⁰.

Para tanto, os autores propõem um esquema (na Ilustração 13) para ilustrar como se dá a conversão do fonema para o som, i.e., trata-se de um esquema de produção da fala que, devido às características apresentadas, possibilita ser utilizado também na via da percepção.

Vale a ressalva aqui de que os autores estão assumindo que a percepção é mediada pela produção ao invés da audição. O argumento deles é baseado nos achados de células do córtex auditivo de gatos que são sensíveis à frequência. Eles apontam que, embora essas células também sejam encontradas no sistema nervoso humano, não há comprovação de como essas células se comportam com sinais mais complexos e de como se daria conta de transições diferentes que se ligam a um mesmo fonema.

Na caixa mais ao topo da ilustração 13 há as regras sintáticas (*Syntactic Rules*) que podem ser divididas em regras de estrutura frasal, transformacionais e morfofonêmicas⁶¹. Os autores não se detêm aos detalhes do modelo sintático, pois nesse momento apenas estão preocupados com a percepção em um sentido mais amplo. Assim, os pesquisadores vão iniciar a proposta com uma mensagem na forma de uma sequência ordenada de fonemas com seus traços constituintes para

⁶⁰ “It seems unparcimonious to assume that the speaker-listener employs two entirely separate processes of equal status, one for encoding language and the other for decoding it. A simpler assumption is that there is only one process, with appropriate linkages between sensory and motor components” (Liberman e colegas, 1967: p.221).

⁶¹ Os autores seguem o modelo sintático presente em Chomsky (1965).

seguir com as conversões necessárias para chegar à forma de onda acústica.

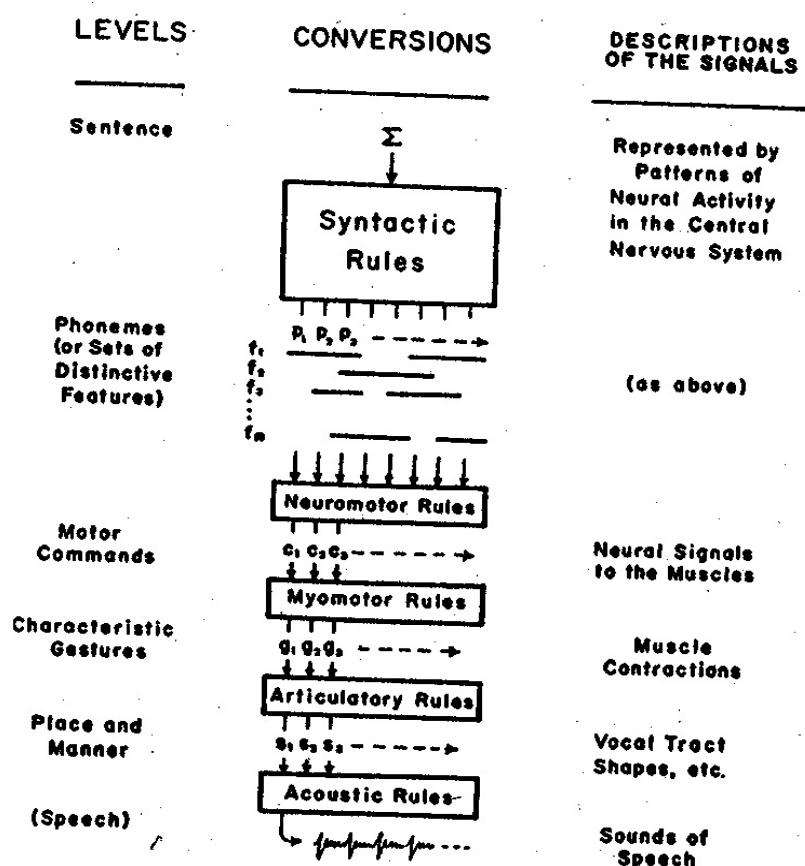


Ilustração 13: Representação esquemática dos estágios assumidos para a produção da fala. (Lieberman e colegas, 1967: p.216).

O modelo é bem simples: as sentenças são apresentadas como sequências ordenadas de fonemas que apresentam constituintes subfonêmicos, chamados aqui de traços. Os autores argumentam que há no sistema nervoso central instruções implícitas que separam a máquina motora em estágios independentes. Assim, os sinais neurais estabelecidos de um para um pelos traços subfonêmicos aos músculos são ajustados em termos de amplitude e de coordenação temporal pelas Regras Neuromotoras (*Neuromotor Rules*). A próxima conversão ocorre do comando neural para a contração do músculo, que é realizada pelas Regras Miomotoras (*Myomotor Rules*). Tais contrações podem ser observadas por eletromiografia (EMG) e a partir delas podemos prever os sinais neurais do próximo nível. No estágio

seguinte, as contrações musculares são convertidas nos formatos que o trato vocal assume. Tais formatos são garantidos pelas Regras Articulatorias (*Articulatory Rules*) e é aqui que a reestruturação do fonema para a articulação (e depois para o som) ocorre. Embora seja previsível, trata-se sobretudo de uma relação complexa, pois

“a verdadeira codificação ocorre como consequência de dois outros aspectos nesta conversão; o fato de que os traços subfonêmicos podem ser, e são, colocados em paralelo significa que cada novo conjunto de contrações (a) começa de qualquer configuração que já exista (como resultado de uma contração precedente) e (b) tipicamente ocorre antes que o último conjunto tenha terminado, com o resultado de que o formato do trato em qualquer instante represente os efeitos provenientes das instruções passadas e presentes” (Liberman e colegas, 1967: p.217)⁶².

Como podemos perceber, a relação entre a mensagem e o código vai se tornando cada vez mais complexa à medida que as sobreposições temporais e espaciais ocorrem simultaneamente. Diante disso, a conversão das contrações musculares para os formatos que o trato vocal assume é suficiente para mostrar como ocorre a perda da segmentalidade e das mudanças acústicas em função do contexto, i.e., “dada a estrutura do aparato articulatório, essas complexidades parecem ser uma necessidade concomitante ao processamento em paralelo que faz do código da fala tão eficiente” (Liberman e colegas, 1967: p.217)

Por fim, a última conversão é referente ao formato do trato vocal para o som e ela é representada pelas Regras Acústicas (*Acoustic Rules*). Essas regras operam a partir de uma relação de um para um entre o formato para o som, garantindo assim que as diferenças acústicas provenientes dos contextos sejam produzidas e passíveis de serem percebidas no “caminho de volta” do som para o fonema.

Neste momento, o texto chega a seu ponto central: a decodificação da fala só é possível através de um decodificador especial. Mais uma vez, o argumento reside na falta de correspondência entre o acústico e o articulatório e no fato de que, se

⁶² “True encoding occurs as a consequence of two further aspects of this conversion; the fact that the submorphemic features can be, and are, put through in parallel means that each new set of contractions (a) starts from whatever configuration then exists (as the result of the preceding contractions) and (b) typically occurs before the last set has ended, with the result that the shape of the tract at any instant represents the merged effects of past and present instructions” (Liberman e colegas, 1967: p.217).

instâncias superiores aos fonemas (como as sílabas) forem requisitadas para se ligarem às inúmeras variações acústicas, teremos um número imensamente grande de possibilidades de configurações silábicas. Em outras palavras,

“perceber a fala pela correspondência entre o sinal acústico para muitos padrões pareceria deselegante e não econômico. Ainda mais importante, isso seria com certeza inadequado, uma vez que isso deve falhar para dar conta do fato de que percebemos fonemas. Nenhuma teoria pode ignorar seguramente o dado de que os fonemas são psicologicamente reais” (Liberman e colegas, 1967: p.220)⁶³.

Conforme apresentado rapidamente em parágrafos anteriores, os autores vão sugerir que o processo de percepção é mediado pela produção sob a justificativa de que se trata de um modelo mais parcimonioso.

À parte disso, há ainda outros argumentos que endossam a proposta de que “a percepção acompanha mais a articulação do que o som” (Liberman e colegas, 1967: p.221). O primeiro é o já supra citado caso de que o /d/ é percebido como /d/ mesmo não havendo variação acústica a depender do contexto adjacente (/di/ e /du/). Em um outro trabalho, Liberman, Delattre & Cooper (1952) relatam que essa falta de correspondência fica ainda mais complexa à medida que se considera o comportamento de outras consoantes. Por exemplo, um mesmo *burst* é percebido como /p/ antes de /i/, mas como /k/ antes de /a/. A explicação deles é a de que há diferenças na sobreposição temporal e, conseqüentemente, na codificação acústica, uma vez que parecem ser necessários diferentes ambientes vocálicos para produzir um mesmo efeito acústico (o *burst* da consoante). Um outro exemplo é o relacionado à consoante /g/, que apresenta características bem distintas, mesmo em contexto vocálico parecido, como nas diferenças entre /ga/ e /gɔ/ (cf. Liberman 1957 *apud* Liberman e colegas, 1967: p.221). Com esse achado, poder-se-ia argumentar que os falantes utilizam sinais acústicos constantes e fazem uso de mudanças drásticas na articulação. No entanto, não fazemos isso, pois ainda ouvimos o mesmo fonema. Tais fatos suportam a assunção de que os ouvintes usam um sinal inconstante como base para encontrar o caminho de volta para os gestos articulatórios produzidos e,

⁶³ “To perceive speech by matching the acoustic signal to só many patterns would appear, at best, uneconomical and inelegant. More important, it would surely be inadequate, since it must fail to account for fact that we do perceive phonemes. No theory can safely ignore the fact tha phonemes are psychologically real” (Liberman e colegas, 1967: p. 220).

no limite, o que o falante pretendeu⁶⁴. Por fim, os dados de percepção categórica também endossam essa proposta, uma vez que a descontinuidade da percepção, mesmo com estímulos distribuídos em um contínuo sonoro, sugere que a percepção prefere seguir a articulação ao sinal acústico.

Em suma,

“a percepção consiste em rodar o processo ao contrário, os sinais neurais correspondem aos vários segmentos que são encontrados nos seus respectivos níveis. Na percepção do fonema – nossa preocupação primária no trabalho – o invariante é encontrado no sistema neuromotor, no nível dos comandos musculares” (Liberman e colegas, 1967: p.222)⁶⁵.

À guisa de conclusão sobre o processo de decodificação, os autores ainda salientam que, como consequência do fato de os comandos serem os invariantes para os diferentes formatos acústicos assumidos e percebidos como um único fonema (como é o caso de /d/), mesmo os ouvintes mais cuidadosos não têm consciência de que os sinais acústicos são diferentes. Em contrapartida, o falante sabe que o “morfofonema” /S/ é um só tanto nas produções como /s/ em “cats” quanto /z/ em “dogs”⁶⁶, i.e., “ele sabe que esses dois formatos acústicos e fonéticos são em certo sentido o mesmo” (Liberman e colegas, 1967: p.223).

Para os autores, tais argumentos parecem ser suficientes para propor que o caminho de volta do som para o fonema só é possível se a percepção for mediada pela produção, por um módulo especial para a fala.

2.3 A três passos da teoria

Antes de passarmos para a proposta mais conhecida da TM, são necessários

⁶⁴ Aqui aparece ainda de forma não sistematizada a noção de que se percebem os gestos pretendidos. Como veremos nas seções seguintes (2.3.1 e 2.3.3), essa assunção só é desenvolvida com mais propriedade após o advento da Fonologia Gerativa de Chomsky & Halle (1968) e a proposta da Modularidade da Mente de Jerry Fodor (1983).

⁶⁵ “Perception consists in somehow running the process backward, the neural signals corresponding to the various segments being found at their respective levels. In phoneme perception – our primary concern in this paper – the invariant is found far down in the neuromotor system, at the level of the commands to the muscles” (Liberman e colegas, 1967: p.222).

⁶⁶ O exemplo dado pelos autores vincula escancaradamente a sua filiação a uma fonologia estruturalista (ou fonêmica), pois há a noção não só de fonema (que virá a ser desconsiderada no gerativismo), como também a de arquifonema. E, diga-se de passagem, a noção de Troubetzkoy de que a “fonologia é o que se crê pronunciar”.

parênteses para esclarecer onde ela se fundamenta. Nesta seção, veremos que o advento da Fonologia Gerativa de Chomsky & Halle (1968), os achados referentes ao Efeito McGurk (McDonald & McGurk, 1976) e a proposta da Modularidade da Mente de Jerry Fodor (1983) são o *pulo do gato* para uma versão mais forte da Teoria Motora da Percepção da Fala.

2.3.1 A Fonologia Gerativa: *The Sound Pattern of English* (Chomsky & Halle 1968)

Nesta seção, apresentaremos, em linhas gerais, a proposta de organização da fonologia dentro do arcabouço gerativo transformacional para, em seguida, localizar quais dessas informações foram incorporadas na versão revisada da TM.

Antes de tudo, é preciso ressaltar que “*The sound pattern of English*” (SPE), de Chomsky & Halle (1968), pode ser considerada um ponto importante nos estudos de fonologia, devido à sua vinculação ao revolucionário paradigma gerativista transformacional e à inclusão de informações fonéticas no nível abstrato da fonologia. Tamanha importância pode ser verificada pela constituição de duas gerações de foneticistas e fonólogos após a sua publicação. Albano (2001: p.36-37) faz a seguinte constatação:

“A primeira geração é marcada pelo estruturalismo e fez uma leitura um tanto cética de SPE, pautando-se antes pela orientação de *Preliminaries* de buscar os correlatos auditivos e motores do traço para deduzir os primitivos da Fonologia de restrições biológicas mais gerais. Os representantes que nos interessarão mais de perto são John Ohala, Alvin Liberman, Björn Lindblom e Kenneth Stevens. [...] A segunda geração levou a sério o esquema da Figura 1.1⁶⁷ e vem trabalhando as consequências da sua refutação. Aceitou o ônus da prova quanto à existência de alofonias contínuas e estabeleceu que não existe uma mecânica verdadeiramente ‘intrínseca’ da produção da fala, sem influência da língua que está sendo exercitada. Constitui certamente a ponta atual do campo, embora, por razões, conjunturais, só desfrute de reconhecimento nos meios simpatizantes. A precisão com que focaliza e aprofunda questões específicas deve-se, sem dúvida, à eficácia da limpeza geral de terreno empreendida pela geração anterior”.

⁶⁷ A figura 1.1. (disponível em Albano 2001: p.18) é reproduzida como ilustração 14 neste trabalho.

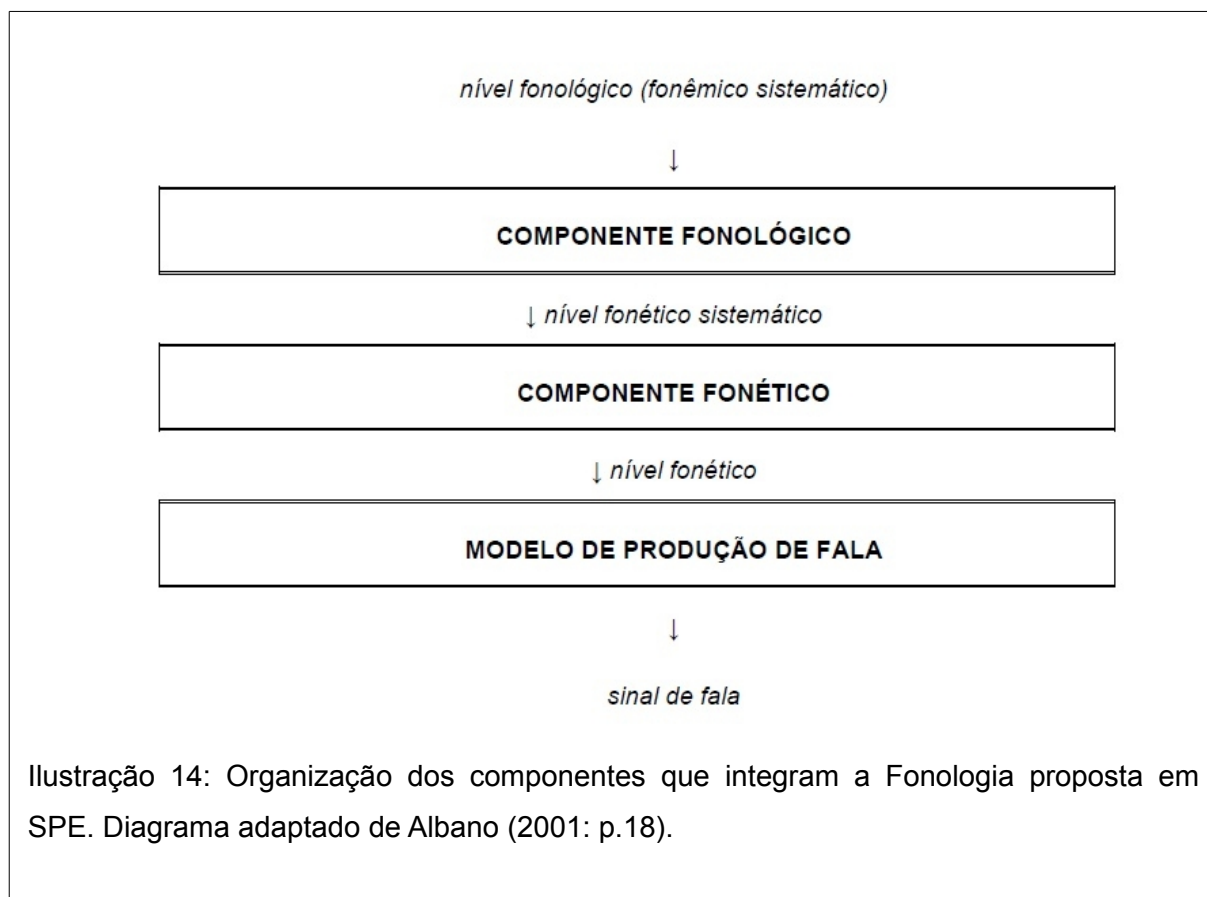
Em linhas gerais, podemos dizer que o frescor da proposta de SPE reside justamente nos conceitos abordados dentro do arcabouço gerativista. Ao contrário das tentativas de se buscar os traços distintivos subjacentes aos fonemas durante o predomínio da Fonologia Estruturalista, no SPE adota-se um outro olhar sobre o objeto da fonologia: trata-se de se buscar quais são as regras universais e específicas de língua que determinam a correspondência entre som e significado realizada pelos falantes e ouvintes de uma língua.

Para delimitar exatamente o objeto de análise do modelo, os autores lançam mão da dicotomia *desempenho* e *competência*. Tal qual nos estudos em sintaxe, a fonologia vai estudar o conhecimento que o falante tem sobre a gramática que determina a conexão intrínseca entre som e significado de cada sentença. O conhecimento “inconsciente” do usuário da língua é a competência e é ela que deve ser estudada. Em suas palavras: “nós podemos, se quisermos, pensar o estudo da competência como o estudo da performance potencial de um falante-ouvinte idealizado que não é afetado por tais fatores gramaticalmente irrelevantes” (Chomsky & Halle, 1968: p.3).

Basicamente, a proposta de SPE é a seguinte: a gramática é dividida em componentes. O Componente Sintático vai entregar ao Componente Fonológico uma estrutura superficial que “determina totalmente a forma fonética da sentença” (Chomsky & Halle, 1968: p.6). É neste módulo (o Fonológico Universal) que os enunciados vindos do módulo sintático têm as regras fonológicas específicas de língua aplicadas. Assim, tendo sido realizada a aplicação dessas regras, “uma certa representação fonética esboçada a partir da classe universal fornecida pela teoria linguística geral” (Chomsky & Halle, 1968: p.9) é gerada. Trata-se de uma representação equivalente à transcrição fonética (ainda anterior ao sinal de fala), que é então passada ao “Componente Fonético”. É nesta etapa que se aplicam as regras universais de língua. Ao contrário das regras baseadas em traços binários do componente fonológico, as regras do Componente Fonético são baseadas em escalas físicas, de modo que acabam por integrar detalhes fonéticos finos à representação. Embora as regras sejam baseadas em uma escala física, vale ressaltar que essas especificações são categóricas e abstratas, isto é, não se trata de especificações dispersas em um *continuum* físico. Por fim, depois de recebidas

essas especificações, há o processamento do sinal de fala.

Essa organização por componentes pode ser visualizada a partir do diagrama proposto por Albano (2001: p.18):



Uma outra questão inovadora na proposta de SPE é como são apresentadas tais formas abstratas que são afetadas pelas regras fonológicas, sejam elas universais ou específicas de língua. Ao contrário da proposta Estruturalista, as representações têm “a forma de uma matriz bidimensional na qual as linhas representam traços fonéticos particulares; as colunas representam segmentos consecutivos da sentença gerada; e as entradas na matriz determinam o *status* de cada segmento com respeito aos traços” (Chomsky & Halle, 1968: p.5). Esse tipo de representação não é meramente um novo *design* de se transcreverem as regras, trata-se, sobretudo, de uma postura teórica.

Ao delimitarem os componentes que integram o sistema, os autores

argumentam que evitaram chamá-los de *representação morfofonêmica* ou *representação fonêmica sistemática*, pois, se assim os chamassem, eles estariam assumindo que haveria um nível fonêmico abstrato em algum ponto da derivação. Diante disso, eles acabam fazendo nenhuma menção à *análise fonêmica* ou *fonema*. Eles ainda complementam a questão da seguinte forma:

“Note que, neste caso, a questão não é terminológica mas sim substantiva; a questão é se as regras da gramática devem ser tão limitadas como para fornecer, em algum estágio da geração, um sistema de representação que encontre várias condições propostas” (Chomsky & Halle, 1968: p.11⁶⁸).

Em outras palavras, os autores estão aqui propondo que não se utilize mais o fonema como primitivo de análise fonológica. Devem-se, antes de tudo, contemplar os traços (aqui de base articulatória) organizados em matrizes bidimensionais.

Na seção 2.4. (do Capítulo 2), apresentamos a proposta revisada da TM e, lá, é possível verificar que a não utilização do fonema como primitivo de análise é um ganho para a teoria, pois são contemplados os comandos pretendidos pelos falantes e não os fonemas em si (tal como nas versões anteriores do modelo). Tais comandos podem estar vinculados aos traços que integram a matriz de traços de um dado segmento e, como se trata dos comandos pretendidos, abstratos, é inevitável a sua vinculação com as formas abstratas da organização da estrutura fonológica proposta no SPE.

2.3.2 Ouvindo lábios e vendo vozes

McDonald & McGurk (1976) apresentam os resultados de um experimento que tem por objetivo avaliar a percepção da fala a partir de estímulos visuais e acústicos. Tal avaliação surge da contestação sobre a concepção corrente de que a percepção é um processo unicamente auditivo. No entanto, os autores salientam que a comunicação verbal ocorre, na maioria das vezes, em contextos em que o ouvinte vê e ouve simultaneamente seu interlocutor. Diante disso, faz-se necessário avaliar qual seria o “papel” da visão para a percepção da fala, i.e., “o estudo

⁶⁸ “Notice that the issue in this case is not terminological but rather substantive; the issue is whether the rules of a grammar must be so constrained as to provide, at a certain stage of generation, a system of representation meeting various proposed conditions” (Chomsky & Halle, 1968: p.11).

apresentado aqui demonstra uma influência da visão não reconhecida anteriormente sobre a percepção da fala” (McDonald & McGurk, 1976: p.746)⁶⁹.

Seu experimento parte da observação de que ao se mostrar a imagem (em um vídeo) do rosto de uma mulher articulando a sílaba [ga], mas com o audio dublado produzindo a sílaba [ba], adultos ouvem a [da]. No entanto, ao se apenas ouvir o som (sem levar em conta a imagem), os sujeitos envolvidos no experimento ouvem o som [ba]. Para os autores, esse fenômeno (que depois ficou conhecido como *Efeito McGurk*⁷⁰) “tem implicações importantes para o entendimento da percepção da fala” (McDonald & McGurk, 1976: p.746) e, por isso, fez-se necessário replicá-lo a fim de se obter confirmações e generalizações mais robustas.

O experimento consiste em pedir que os informantes, em um primeiro momento, ouçam e repitam os estímulos sonoros apresentados sem a utilização da imagem de alguém articulando os sons. Trata-se da etapa auditiva do experimento. Em um segundo momento (a etapa visual-auditiva), os informantes são solicitados a ouvir e repetir os estímulos sonoros com a utilização da imagem de uma pessoa articulando os sons.

O que há de novo (e que promove um novo olhar sobre a percepção da fala) é a utilização de estímulos visuais e auditivos “trocados”. A partir da gravação da imagem e do audio de uma pessoa produzindo as sílabas [ba]-[ba], [ga]-[ga], [pa]-[pa] e [ka]-[ka], os experimentadores geraram 4 vídeos dublados: 1) [ba]-voz/[ga]-lábios; 2) [ga]-voz/[ba]-lábios; 3) [pa]-voz/[ka]-lábios; e 4) [ka]-voz/[pa]-lábios. Os informantes foram divididos em 3 grupos: 1) 21 crianças em fase pré-escolar (entre 2 e 4 anos); 2) 28 crianças já em fase escolar (entre 7 e 8 anos); e 3) 54 adultos (entre 18 e 40 anos). Todos eles foram submetidos aos dados aleatorizados e sob as duas condições já mencionadas, a saber: 1) experimento somente com base na audição (etapa auditiva: ouvir e repetir); e 2) experimento com os estímulos visuais e auditivos (ver, ouvir e repetir).

Os resultados da etapa auditiva mostram que todos os informantes do experimento apresentaram uma taxa de acerto alta. Para o grupo das crianças em

⁶⁹ “The study reported here demonstrates a previously unrecognised influence of vision upon speech” (McDonald & McGurk, 1976: p.746).

⁷⁰ Para experienciar como o efeito ocorre, visite: <http://www.youtube.com/watch?v=jtsfidRq2tw>.

fase pré-escolar a taxa foi de 91%, enquanto os outros dois grupos obtiveram acerto quase de 100%. As crianças em fase escolar acertaram 97% e os adultos, 99%. Por sua vez, as respostas dos informantes na etapa visual-auditiva não seguiram a tendência de acertos da etapa anterior. O grupo das crianças em fase pré-escolar obteve taxa de erro de 59% e o das em fase escolar, 52%. Mais interessante ainda é a taxa de erro dos adultos que chega à marca de 92% de erro.

Ao se analisar os resultados dessa etapa do experimento, os autores notam que há dois tipos de respostas (Ilustração 15). Os autores as chamam de “fundidas” e “combinadas”. As chamadas respostas “fundidas” são aquelas que mostram que ocorreu o *Efeito McGurk*, i.e., os informantes não perceberam apenas o componente auditivo ou o visual. Eles perceberam uma terceira possibilidade que é [da], ao invés de [ba] ou [ga]. Em outras palavras, os participantes do experimento transformaram as duas informações disponíveis (visual e auditiva) em algo novo que não estava presente nos estímulos.

Já as respostas “combinadas” indicam que o *efeito* não ocorreu, pois há a percepção “misturada” dos estímulos visuais e sonoros. Notem-se que as respostas são do tipo [gabga], [bagba], [baga] ou [gaba] (bem como as suas contrapartes surdas). Trata-se, pois, de uma composição comprimida (e sem modificações) dos elementos presentes em cada uma das modalidades de estimulação sensorial.

Table 1 Stimulus conditions and definition of response categories from auditory-visual condition						
Stimuli		Response Categories				
Auditory component	Visual component	Auditory	Visual	Fused	Combination	Other
ba-ba	ga-ga	ba-ba	ga-ga	da-da	—	—
ga-ga	ba-ba	ga-ga	ba-ba	da-da	gabga bagba baga gaba	dabda gagla etc.
pa-pa	ka-ka	pa-pa	ka-ka	ta-ta	—	tapa pla kafta etc.
ka-ka	pa-pa	ka-ka	pa-pa		kapka pakpa puka kapa	kat kafa kakpat etc.

Ilustração 15: Tabela presente em McDonald & McGurk (1976: p.747) que apresenta as respostas obtidas em seu experimento.

Na Ilustração 16, é possível verificar as porcentagens das respostas dadas pelos informantes em cada uma das condições do experimento. Inicialmente, o que se interpreta dessa imagem é que o *efeito* se comprova com os adultos nos padrões 1 e 3, i.e., o *efeito* ocorre quando há um estímulo sonoro bilabial é sobreposto ao visual que articula uma velar ([ba]-voz/[ga]-lábios e [pa]-voz/[ka]-lábios). Segundo os autores, isso mostra que a percepção auditiva dos adultos é mais influenciada pelos aspectos visuais envolvidos na produção da fala do que para os grupos mais jovens. A ilustração 16 ainda mostra que nos padrões 2 e 4 pode ocorrer o efeito, mas há um número maior de respostas “combinadas” do que “fundidas”.

Stimuli		Subjects	Responses		Responses Fused	Combination	Other
Auditory	Visual		Auditory	Visual			
ba-ba	ga-ga	3-5 yr (n=21)	19	0	81	0	0
		7-8 yr (n=28)	36	0	64	0	0
		18-40 yr (n=54)	2	0	98	0	0
ga-ga	ba-ba	3-5 yr (n=21)	57	10	0	19	14
		7-8 yr (n=28)	36	21	11	32	0
		18-40 yr (n=54)	11	31	0	54	4
pa-pa	ka-ka	3-5 yr (n=21)	24	0	52	0	24
		7-8 yr (n=28)	50	0	50	0	0
		18-40 yr (n=54)	6	7	81	0	6
ka-ka	pa-pa	3-5 yr (n=21)	62	9	0	5	24
		7-8 yr (n=28)	68	0	0	32	0
		18-40 yr (n=54)	13	37	0	44	6

Ilustração 16: Tabela presente em McDonald & McGurk (1976: p.747) que apresenta as porcentagens das respostas dadas pelos informantes em cada uma das condições do experimento.

Esses dados apresentados por McDonald & McGurk (1976) são interessantes devido ao fato de eles comentarem que o simples fato de se fechar os olhos durante a experiência de base auditivo-visual ocorre a mudança na percepção. Por exemplo, ao se experimentar o *efeito*, um sujeito ouve [da] enquanto há um som [ba] sincronizado com uma imagem articulando [ga]. Entretanto, ao se fechar os olhos, imediatamente o sujeito passará a ouvir [ba] apenas. Para os autores,

“nos dias atuais, teorias de percepção da fala de base auditiva são inadequadas para acomodar essas novas observações; um papel da visão (ou seja, os movimentos dos lábios percebidos) na percepção da fala por pessoas que ouvem normalmente é claramente ilustrado”

(McDonald & McGurk, 1976: p.747)⁷¹.

Após terem claramente criticado as abordagens de base unicamente auditivas, os autores tentam dar explicações levando em conta tanto os aspectos sonoros quanto visuais para explicar o fato de ocorrerem percepções “fundidas” e “combinadas”. Primeiramente, eles comentam que (cf. Binnie, Montgomery & Jackman, 1974 *apud* McDonald & McGurk, 1976: p.747) na ausência de estímulos auditivos normalmente se ouve um [ga] como [da] e um [ka] como [ta]. Por sua vez, [ba] e [pa] são constantemente confundidos com outros sons se não se levarem em conta os aspectos visuais, uma vez que, na presença dos estímulos visuais, eles nunca são confundidos com [ga, da, ka, ta].

Diante disso e apoiando-se nos trabalhos de Liberman, Delattre & Cooper (1952) que argumentam que as vogais carregam informações acústicas sobre as consoantes adjacentes a elas, McDonald & McGurk especulam uma explicação para ocorrer o *efeito*: ele somente acontece porque há traços acústicos compartilhadas entre [ba] e [da] e visuais entre [da] e [ga]. Assim, estando o sujeito diante de informações acústicas que apontam para um [ba] ou [da] e de estímulos visuais que podem ser um [da] ou [ga], eles acaba percebendo [da] por cruzar as informações nos dois níveis de estimulação, unificando os estímulos disponíveis.

Para os autores, essa especulação também explicaria as respostas combinadas, pois como não haveria traços compartilhados pelos dois níveis sensoriais, os sujeitos não teriam como decidir sobre a fonte de informação. Deste modo, ele criaria *percepts* que apresentariam todas as informações disponíveis, tais como [babga], [baga] ou [papka].

Como se pode ver, os dados apresentados por McDonald & McGurk (1976) ainda não possuíam um tratamento de base articulatória. Eles apenas apontavam para a importância dos estímulos visuais para a percepção da fala por pessoas sem problemas de audição. Esses dados serão incorporados como evidências experimentais para a percepção de gestos pretendidos na versão revisada da TM (Liberman & Mattingly, 1985). Antes de passarmos à versão revisada da teoria, apresentaremos um outro pilar teórico que sustenta a TM. Trata-se da proposta

⁷¹ “Contemporary, auditory-based theories of speech perception are inadequate to accomodate these new observations; a role for vision (that is, perceived lip movements) in the perception of speech by normally hearing people is clearly illustrated” (McDonald & McGurk, 1976: p.747).

modular sobre a mente de Fodor (1983). Vamos a ela.

2.3.3 A base da Teoria Motora: “A Modularidade da Mente” de Jerry Fodor (1983)

Na introdução de *A Modularidade da Mente* (MM), Fodor (1983) argumenta que há uma retomada nos estudos sobre a psicologia das faculdades mentais. Entende-se por psicologia das faculdades mentais uma maneira de encarar e postular mecanismos diferentes para se explicar a heterogeneidade da vida mental dos seres humanos, i.e., trata-se de levar a sério a diversidade da mente e, assim, tentar estabelecer as diferenças entre a sensação e a percepção, a volição e a cognição, aprendizagem e lembrança ou a linguagem e o pensamento. Tal retorno, salienta Candiottto (2008), é uma tentativa de negar as abordagens horizontalistas sobre a mente e, com isso, fomentar as abordagens verticalistas.

O programa de pesquisa de Fodor consiste em estudar cada uma das características implícitas nas faculdades citadas e buscar entender como elas interagem. Para tanto ele estabelece alguns pontos a serem desenvolvidos na sua tese, a saber: (i) distinguir o argumento geral das faculdades psicológicas de um particular: a tese da modularidade; (ii) enumerar algumas propriedades que um sistema cognitivo modular parece exibir em virtude da sua modularidade; e (iii) “desvincular” a tese da modularidade da tese da Limitação Epistêmica, i.e., “a ideia de que há restrições determinadas endogenamente nos tipos de problemas que os seres humanos podem resolver, por conseguinte os tipos de coisas que nós podemos saber” (Fodor, 1983: p.2). Vamos a elas!

2.3.3.1 Quatro tratamentos para a estrutura mental

Fodor apresenta quatro tratamentos possíveis para a estrutura mental, são elas: neocartesianismo; horizontalismo; verticalismo; e associacionismo. Ele apresenta tais propostas para que fique claro em quais pontos a sua tese difere e se iguala às demais. Para ele, é das diferenças entre essas abordagens que surge o

argumento para a adoção de uma visão modular da mente que, em suas palavras, é “descaradamente eclética” (Fodor, 1983: p. 3).

Tendo como base o trabalho do filósofo René Descartes, o neocartesianismo tem como principal expoente Noam Chomsky. Sua tese é a de que “a mente é (inicial, intrínseca e geneticamente) estruturada em faculdades psicológicas ou em órgãos” (Fodor, 1983: p.3), i.e., para ele, a mente é entendida como órgãos da mesma maneira que o coração e os pulmões, por exemplo. Essa assunção assume tacitamente que há um “processo determinado” no desenvolvimento da mente.

Mais especificamente, o que é inato é um conjunto de informações (com conteúdo proposicional) sobre “restrições universais da fala das línguas humanas possíveis” (p. 4) e, tal conhecimento inato, integrado com dados linguísticos primários, dá as características da sua fala adulta à fala dos aprendizes de uma dada língua.

Para Fodor,

“quando Chomsky diz que há um 'órgão de linguagem' especificado inatamente, significa dizer primeiramente que há verdades (sobre a estrutura das possíveis primeiras línguas) que os seres humanos compreendem inatamente. Quando ele diz que a mente de uma criança é 'intrinsecamente estruturada', significa primeiramente que há conteúdos proposicionais especificados inatamente. Quando ele diz que a teoria de aprendizagem da linguagem é a história de como a faculdade da linguagem matura, significa primeiramente que a ontogenia das capacidades linguísticas é o desdobramento das consequências dedutivas das crenças inatas em interação com um conjunto de dados perceptuais” (Fodor, 1983: p.7)⁷².

Diante disso, Fodor vai aproveitar da abordagem chomskyana a noção de que a estrutura modular da mente e os seus respectivos mecanismos psicológicos são inatos, i.e., ao invés de o conteúdo dos módulos ser proposições computadas via transformação, o autor conta com as faculdades psicológicas para mediar “a

⁷² “[...], when Chomsky says that there is an innately specified 'language organ', what he means is primarily that there are truths (about the structure of possible first languages) that human beings innately grasp. When he says that the mind of the child is 'intrinsically structured', what he means is primarily that there are innately specified propositional contents. When he says that the theory of language learning is the story of how the language faculty matures, what he means is primarily the the ontogeny of linguistic capacities is the unfolding of the deductive consequences of the innate beliefs in interaction with a body of perceptual data” (Fodor, 1983: p.7).

representação, a retenção, a recuperação e elaboração referencial das proposições cognitivizadas” (p.10). Ou seja, a estrutura mental não é composta por representações internas de conteúdos proposicionais, a ponto de essas estruturas inatas não poderem ser uma teoria de crenças inatas. Há uma diferença que é obscurecida ao se apropriar da terminologia das faculdades mentais, órgãos e mecanismos para expressar “um nativismo de atitudes proposicionais” (p.10) e não um nativismo de representações proposicionais.

Tais argumentos de Fodor são o que ele chamou de pontos persuasivos da teoria de chomskyana. Por exemplo, o autor de “A Modularidade da Mente” levanta a questão relativa à possibilidade de considerar a memória inata. Para ele, ela seria um candidato plausível para ser especificado inatamente. No entanto, a memória não seria uma faculdade no sentido neocartesiano, i.e., ela não poderia ser uma capacidade inata dessa natureza porque há conjuntos de proposições que os seres nascem sabendo e “não há a mais remota tentação de identificar a estrutura da memória com uma estrutura inferencial de um conjunto de proposições” (p. 8). Para ele, aproveitando-se também das metáfora anatômicas, a memória pode ser um mecanismo como braços e coração, por exemplo, mas apenas um mecanismo e não um órgão.

Por exemplo, Fodor cita a tese de George Miller sobre as propriedades mágicas do número sete (cf. Miller, 1956 *apud* Fodor, 1983: p.8). O aspecto mágico desse número faz com que uma pessoa, ao ser solicitada repentinamente a realizar uma determinada tarefa, só consiga realizar um número limitado de atividades. Assim, ao solicitarmos que alguém dê uma lista de palavras *nonsense*, ela conseguirá produzir, em média, uma lista com sete palavras.

Uma maneira neocartesiana de lidar com esses fatos é postular que há uma proposição inata que diz “*quando apresentado a uma lista de n coisas a aprender, a pessoa deve aprender as sete primeiras e não se importar com o resto*”. Embora essa proposição não seja “levada a sério” por neocartesianos, esta seria uma maneira de tratar do caso. No entanto, o que ocorre parece ser muito mais um mecanismo psicológico que impõe limitações sobre as capacidades do sistema, ao invés de ser uma regra inatamente postulada na cognição. Como se pode ver, há diferenças de tratamento se se postula uma noção tipicamente neocartesiana (de

conteúdo proposicional) e outra que toma por base mecanismos psicológicos. É sobre esta que Fodor vai estabelecer sua proposta.

Uma segunda maneira de estruturar a mente é considerá-la uma arquitetura funcional de faculdades horizontais, i.e., enquanto os neocartesianos estruturam a mente com relação aos conteúdos proposicionais típicos, na proposta horizontalista a faculdade é individualizada pela “referência a seus efeitos típicos, o que significa dizer que é individualizado funcionalmente” (p.10).

Em suma, Fodor apresenta a versão mais conhecida do que é uma faculdade psicológica horizontal:

“[...] os processos cognitivos exibem a interação de tais faculdades como, por exemplo, memória, imaginação, atenção, sensibilidade, percepção, e assim por diante; e a característica de cada um desses processos é determinada pela mescla particular das faculdades que ela recruta. Entretanto, a característica da mentalidade é mais ou menos independente da questão do sujeito; as faculdades devem ser invariantes de um tópico de pensamento para outro” (Fodor, 1983: p.11)⁷³.

Isso significa dizer que uma faculdade se estende por mais de um domínio cognitivo, exercitando-se em relação a distintos conteúdos. Por exemplo, o tratamento tradicional sobre a memória na literatura diz que se trata de um lugar onde as crenças são estocadas⁷⁴. Essa analogia é horizontalista, pois se pressupõe que todas as memórias estão em um mesmo lugar e, mesmo se estiverem em um único lugar e separadas por pequenos sistemas, “cada memória deve passar presumivelmente por todo o sistema” (p.12) quando quiser acessar alguma, i.e., sejam as memórias sobre as cores de uma flor ou sobre os eventos ocorridos em um parque de diversões, estão todos disponíveis a todos os processos cognitivos.

Essa abordagem pode ser chamada de espacial, porque se fala de uma localização do conteúdo mental em um dado ponto *t* do tempo. Devido ao problema de assumir uma proposta dessa natureza, ganha força uma abordagem funcional das faculdades horizontais. Trata-se de um conjunto de processos mentais que têm

⁷³ “[...] cognitive process exhibit the interaction of such faculties as, e.g., memory, imagination, attention, sensibility, perception, and so forth; and the character of each such process is determined by the particular mix of faculties that it recruits. However, the character of mentation is more or less independent of its subject matter; the faculties are supposed to be invariant from one topic of thought to the next” (Fodor, 1983: p.11).

⁷⁴ Fodor cita a metáfora da gaiola de Platão (p.12).

acesso a tal conteúdo em *t*. Em outras palavras, um conjunto de processos constituem um domínio em *t*. Assim, “uma faculdade horizontal de modo completo, individualizada funcionalmente, é portanto aquela que todo o conteúdo mental pode estar acessível a qualquer tempo” (Fodor, 1983: p.13)⁷⁵.

Dessa proposta, Fodor vai aproveitar o fato de que os processos mentais são computacionais e cognitivos, uma vez que “a função típica dos mecanismos cognitivos é a transformação das representações mentais” e “disto decorre que cada faculdade cognitiva deve afetar um padrão característico de tais transformações” (cf, Fodor, 1975 *apud* Fodor, 1983: p.13)⁷⁶.

Uma outra maneira de estruturar a mente é a partir de faculdades verticais. Sua origem se encontra no trabalho de Gall (1858-1928) que postula que “não há faculdades horizontais” (p.14). Segundo Fodor,

“ele [Gall] concorda com as teorias tradicionais de faculdade que a mente é estruturada em subsistemas distinguíveis funcionalmente, mas não concorda sobre como as divisões entre esses sistemas seriam desenhadas” (Fodor, 1983: p.14)⁷⁷.

O cerne da proposta de Gall está em dois pontos: (i) no fato de que as aptidões são passíveis de identificação em certos domínios cognitivos, a chamada *especificidade de domínio*; e (ii) na determinação genética dos mecanismos neurais que subservem às faculdades verticais.

Com relação ao domínio específico, “Gall também argumenta que os mecanismos psicológicos que subservem uma capacidade são diferentes, de fato, daqueles que subservem outra” (p.15), i.e., as faculdades verticais “correspondem a mecanismos específicos no cérebro que Gall esperava, cedo ou tarde, ” (p.16). Por exemplo, Gall nega a existência de acuidade. É claro que existe acuidade na visão, audição e tato, por exemplo. No entanto, é possível que as pessoas tenham acuidade na visão e não na audição. Este fato pode indicar que não exista uma

⁷⁵ “A thoroughly horizontal faculty, functionally individuated, is thus one to which every mental content may be accessible at one time or other” (Fodor, 1983: p.13).

⁷⁶ “I shall assume without argument that mental processes are computational insofar as they are cognitive, hence that the typical function of cognitive mechanisms is the transformation of mental representations (see Fodor, 1975)” (Fodor, 1983: p.13).

⁷⁷ “he agrees with traditional faculty theories that the mind is structured into functionally distinguishable subsystems, but disagrees about how the divisions between these systems should be drawn” (Fodor, 1983: p.14).

faculdade de acuidade, já que se trata de um parâmetro para visão, audição e tato. Diante disso, para Gall, todas as outras faculdades horizontais (memória, julgamento, volição, atenção, etc) não existiriam. Assim, “percepção e memória são apenas atributos comuns para as qualidades psicológicas fundamentais, mas não faculdades por elas mesmas; e, conseqüentemente, elas podem não ter nenhum centro próprio no cérebro” (p.16).

Com relação à determinação genética, Gall enfatiza que a especificidade de domínio dos mecanismos neurais que subservem a uma faculdade é inata. Em suas palavras:

“a influência da educação, instrução, exemplos e as circunstâncias ambientais atuam principalmente quando as disposições inatas não são tão fracas nem tão ativas [...]. As impressões recebidas pelos nossos sentidos a partir das fontes externas não são as origens das nossas aptidões, talentos, sentimentos, instintos e propensões [...]. As propensões e os instintos, as aptidões e os talentos, as habilidades intelectuais e as qualidades morais dos homens e animais são inatas”⁷⁸ (Fodor, 1983: p.16).

O argumento apresentado por Gall contra o horizontalismo é o seguinte: se há apenas uma faculdade de memória, por exemplo, e alguém é bom em lembrar coisas, então a pessoa tem que ser boa em lembrar de tudo. O seu argumento, segundo Fodor, é um tipo de *Lei de Leibnitz*: “a mesma faculdade não pode ser tanto fraca quanto forte, desta forma se acontece da memória matemática ser fraca e a musical robusta, então a memória que medeia a matemática não pode ser a mesma, tal qual a memória que medeia a música” (p.17).

Fodor, na sua empreitada “descaradamente eclética”, vai realizar críticas ao pensamento de Gall para incorporar alguns ingredientes a sua própria tese. Primeiramente, não haveria problema algum em haver empregos distintos da memória. A única assunção a ser feita sobre o fato de haver empregos distintos sobre a memória matemática e musical, por exemplo, é o fato de essas memórias serem distintas, i.e., “não há razão óbvia por que a mesma faculdade não poderia

⁷⁸ De acordo com Fodor, as citações de Gall foram retiradas de Hollander (1920). Original: “The influence of education, instruction, example and of surrounding circumstances acts principally when the innate dispositions are neither too feeble nor too energetic ... The impressions received through our senses from external sources are not the origins of our aptitudes, talents, sentiments, instincts and propensities ... The propensities and instincts, the aptitudes and talents, the intellectual abilities and moral qualities of men and animals are innate. [pp. 250-251]” (Fodor, 1983: p.16).

ser forte em uma aplicação e fraca na outra, tanto que os empregos não são por si só idênticos” (p.17).

Por exemplo, o fato de que os grandes jogadores de xadrez consigam lembrar das distribuições das peças de uma maneira melhor do que conseguem lembrar de outras coisas não garante que exista um tipo específico de memória para xadrez. Pelo contrário, isso implica em um princípio geral de que “alguém lembra do que entende” (p.18), i.e.,

“A habilidade do enxadrista de lembrar onde as peças estão é portanto parte do seu conhecimento de como elas poderiam ter chegado lá. Isso testemunha o fato de que [a memória] desaparece quando as peças são arranjadas de maneira que não fazem sentido (cf. DeGroot, 1965 *apud* Fodor, 1983: p.18)⁷⁹.

Uma segunda crítica às propostas de Gall contra o horizontalismo se refere à confusão relacionada à determinação genética e especificidade de espécie. O fato de uma característica ser determinada geneticamente não significa que ela seja uma especificidade da espécie. Por exemplo, o simples fato de alguém ser dotado de um conjunto de características fisiológicas e motoras perceptuais para jogar futebol⁸⁰ não implica que há uma faculdade de jogar futebol vinculada aos seres humanos, i.e., não há o menor interesse de vincular uma herança genética a uma base neurológica específica, ou a uma história evolutiva ou ao fato de que os genes jogam futebol. Para Fodor, “o que é instintivo é geneticamente determinado, mas o inverso claramente não há como manter” (p.20).

Em resumo, Fodor vai dizer que os argumentos de Gall contra o horizontalismo são menos persuasivos que os a favor do verticalismo. Diante disso, ele sugere que há uma certa abertura para se pensar em um modelo misto de faculdades psicológicas, no qual a arquitetura mental pode ser, não inteiramente, arranjada verticalmente. Assim, da proposta de Gall, ele vai aproveitar para o seu modelo as noções de domínio específico, determinação genética, a associação a estruturas neurais distintas e a autonomia computacional, que seria importante para a performance de uma dada faculdade.

⁷⁹ “The chess player's ability to remember where the pieces are is thus part and parcel of his grasp of how they might have got there. Witness the fact, that it disappears when the pieces are set down in ways that don't make sense (DeGroot, 1965)” (Fodor, 1983: p.18).

⁸⁰ Fodor faz uma analogia com os jogadores de *baseball* da primeira base.

Por fim, a quarta maneira elencada por Fodor de estruturar a mente se refere ao Associacionismo, que

“está relacionado com o argumento de que existem faculdades de uma maneira parecida com que o fenomenalismo está relacionado ao argumento de que existem mesas e cadeiras; você pode tomá-las como incompatíveis, ou você pode considerar o associacionismo como as faculdades existentes mas que elas têm o status de construto fora de algum tipo mais fundamental de entidade” (Fodor, 1983: p.23)⁸¹.

Para Fodor, assumir o associacionismo é negar as abordagens psicologizantes da mente, pois “o associacionismo ofereceu um tratamento da ontogenia dos processos mentais que dispensaram o postulado de uma arquitetura cognitiva inata em que, em resumo, dispensou a necessidade de faculdades” (p.27), sejam elas verticais ou horizontais.

Assim, para ilustrar a diferença de abordagem entre uma teoria associacionista e outra baseada em faculdades psicológicas, ele cita uma afirmação falsa e um argumento ruim presentes na definição de “faculdade psicológica” do Dicionário de Filosofia e Psicologia de Baldwin (1911 *apud* Fodor, 1983: p.23-24):

“Dizer que uma mente individual possui um certo tipo de faculdade é meramente dizer que ela é capaz de certos estados e processos. Mas nós encontramos em muitos dos psicólogos das gerações anteriores a tendência de tratar as faculdades como se elas fossem causas, ou condições reais, dos estados de processos nos quais elas são manifestadas e de falar delas como agências positivas interagindo umas com as outras. Portanto, persistir na decisão involuntária é estar de acordo com uma extraordinária força de vontade ou com a faculdade de vontade. Certos processos mentais no homem são ditos como tendo a sua fonte na faculdade da razão, e outros processos em animais inferiores são explicados pela existência da faculdade de instinto. Esse modo de explicação simulada tem recebido o nome de Faculdade Psicológica. Locke, ao criticar a frase 'livre arbítrio' [freedom of will], exibiu claramente a natureza da falácia envolvida. 'Nós podemos dizer que a faculdade de canto canta, e a faculdade de dança dança, tal como a de desejo escolhe, ou que a de entendimento

⁸¹ “Roughly, associationism is related to the claim that there are faculties in something like the way that phenomenalism is related to the claim that there are tables and chairs; you can take them to be incompatible, or you can read associationism as saying that faculties exist but that they have the status of constructs out of some more fundamental sort of entity” (Fodor, 1983: p.23).

compreende....” (p.23-24)⁸².

A afirmação falsa reside justamente no fato de que não é porque a mente possui certas capacidades que cada uma delas vai ser uma faculdade, ou seja, há muito mais capacidades mentais do que faculdades psicológicas (até mesmo em uma abordagem extremamente inflacionada de faculdades psicológicas). Por exemplo, dentre as nossas capacidades mentais há a capacidade de somar 1 mais 1, 1 mais 2, e assim infinitamente. Todas elas estão relacionadas a uma única faculdade matemática. Fodor salienta que “o sentido de faculdade não é, em resumo, equivalente à enumeração das capacidades da mente. Ao contrário disso, trata-se de uma teoria da estrutura dos mecanismos causais que subjazem às capacidades da mente” (p.24). Por sua vez, o argumento ruim se refere à assunção de Locke sobre as faculdades, pois nenhum adepto das faculdades psicológicas diria que a faculdade de canto canta ou que a de dança dança. Ao invés disso, eles diriam que o *organismo* canta, dança, escolhe, etc.

Basicamente, Fodor vai elencar algumas noções aceitas pelos associacionistas para explicar a maneira como a mente funciona. Primeiramente, os associacionistas consideram um conjunto de elementos para construir a estrutura psicológica da mente. Para alguns, os reflexos podem ser os elementos que estruturam o comportamento; para outros, as “ideias” são o que estruturam a mente. Segundo, os associacionistas aceitam que há relações de associação definidas entre os reflexos e as ideias e, tais associações, são preservadas por outras associações, i.e., cada associação é produto de outra associação. Terceiro, há o postulado de leis associativas que serão determinadas pela experiência dos organismos. E, quarto, as relações associativas podem receber parâmetros relevantes da estrutura psicológica, uma vez que elas podem ser definidas por

⁸² “To say that an individual mind possesses a certain faculty is merely to say that it is capable of certain states or processes. But we find in many of the earlier psychologists a tendency to treat faculties as if they were causes, or real conditions, of the states or processes in which they are manifested, and to speak of them as positive agencies interacting with each other. Thus persistence in voluntary decision is said to be due to extraordinary strength of will, or to will-power or to the faculty of will. Certain mental processes in man are said to have their source in the faculty of reason, and certain other processes in lower animals are explained by the existence of a faculty of instinct. This mode of pretended explanation has received the name of Faculty Psychology. Locke, in criticizing the phrase ‘freedom of will’, has brought out very clearly the nature of the fallacy involved. ‘We may as properly say that the singing faculty sings, and the dancing faculty dances, as that the will chooses, or that the understanding conceives ...’ (Fodor, 1983: p.23-24).

sua intensidade e os reflexos pelo seu nível operante.

Diante disso, um associacionista, ao ser indagado sobre quantas faculdades há na mente, diz que “se uma faculdade é um mecanismo psicológico primitivo – um poder *fundamental* – então a resposta é: 'apenas uma; apenas a capacidade de formar associações’” (p.29).

Em suma, a proposta associacionista é a de que não haveria faculdades mentais inatas e as associações são mecanismos para gerar redundâncias na mente para propiciar a aprendizagem de comportamentos e de ideias.

Conforme dito anteriormente, Fodor apresenta quatro maneiras de estruturar a mente para propor a sua tese de que a mente é estruturada por módulos cognitivos. Há, contudo, cinco questões que são melhor respondidas se se assumir a hipótese modular. São elas:

- “1) Trata-se de domínio específico? Ou de operações através dos domínios de conteúdo? Isto é, claro, a questão da organização vertical *versus* horizontal; Gall *versus* Platão.
- 2) O sistema computacional é inatamente especificado ou a sua estrutura é formada por um tipo de processo de aprendizagem?
- 3) O sistema computacional é 'montado' (no sentido de colocar tudo junto em um estoque de subprocessos mais elementares) ou sua arquitetura virtual mapeia relativa e diretamente em sua implementação neural?
- 4) Trata-se de ser estruturado⁸³ (no sentido de serem associadas com um sistema neural específico, localizado e elaboradamente estruturado) ou ser implementado por um mecanismo neural relativamente equipotencial?
- 5) Trata-se de ser computacionalmente autônomo (no sentido de Gall) ou de compartilhar recursos horizontais (de memória, atenção, ou o que quer que seja) com outros sistemas cognitivos?” (Fodor, 1983: p.36-37)⁸⁴

⁸³ Fodor utiliza o termo *hardwired*.

⁸⁴ “ 1. Is it domain specific, or do its operations cross content domains? This is, of course, the question of vertical versus horizontal cognitive prganization; Gall versus Plato.
 2. Is the computational system innately specified, or is its structure formed by some sort of learning process?
 3. Is the computational system 'assembled' (in the sense of having been put together from some stock of more elementary subprocesses) or does its virtual architecture map relatively directly onto its neural implementation?
 4. Is it hardwired (in the sense of being associated with specific, localized, and elaborately structured neural systems) or is it implement by relatively equipotencial neural mechanisms?
 5. Is it computationally autonomous (in Gall's sense), or does it share horizontal resources (of memory, attention, or whatever) with other cognitive systems?” (Fodor, 1983: p. 36-37).

Em resumo, a sua proposta é a de que

“os sistemas cognitivos modulares são de domínio específico, especificados inatamente, estruturados, autônomos e não montados. Partindo de que os sistemas modulares são mecanismos computacionais de domínio específico, segue-se que eles são faculdades de espécie vertical” (Fodor, 1983: p.37)⁸⁵.

Na seção seguinte (2.3.3.2 do Capítulo 2), apresentaremos a proposta de Fodor sobre como a mente se organiza em módulos.

2.3.3.2 A organização modular da mente

A proposta de organização modular da mente de Fodor argumenta em favor da existência de mecanismos funcionais na mente dos seres humanos e que tais mecanismos foram organizados por um longo processo de seleção natural.

Basicamente, o autor propõe que a mente é organizada em três sistemas que funcionam sequencialmente. Seriam os sistemas compiladores, os sistemas de entrada e um processador central.

Os sistemas compiladores são os sistemas que dão entrada ao processo cognitivo, ou seja, trata-se de suportes sensoriais que alimentam todo o sistema computacional da mente. Desta forma, fazem parte dos sistemas compiladores a visão e a audição, por exemplo.

Os sistemas de entrada são os módulos da proposta fodoriana. Trata-se de uma operação que interpreta a informação oferecida pelos sistemas compiladores e que a disponibiliza aos processadores centrais.

No entanto, antes de entregar essas informações aos processadores centrais, os módulos realizam operações sobre a informação capturada pelos sentidos, i.e., os analisadores dos sistemas de entrada trabalham por inferência, uma vez que suas premissas de trabalho são as representações das configurações dos estímulos proximais recebidas pelos compiladores e, como resultado, são entregues aos processadores centrais as representações das características dos objetos distais.

⁸⁵ “Roughly, modular cognitive systems are domain specific, innately specified, hardwired, autonomous, and not assembled. Since modular systems are domain-specific computational mechanisms, it follows that they are species of vertical faculties” (Fodor, 1983: p.37).

A proposta de Fodor sobre a existência dos módulos (dos sistemas de entrada) é fundamentada em algumas características que são compartilhadas exclusivamente por esses módulos que, por sua vez, não fazem parte dos processadores centrais.

Primeiro, os sistemas de entrada são de domínio específico (tal como propunha Gall). Seu argumento é baseado nos experimentos conduzidos nos Laboratórios *Haskins*, nos quais se mostram que os sons são percebidos como fala se processados em um domínio específico de linguagem e como não sendo fala se processados fora dele. Este é o caso de se perceber as transições de uma consoante para uma vogal isoladamente ou dentro de uma sentença como sendo *assobios* ou *glissandos* ou como sendo fala, respectivamente. Esses dados mostram que deve haver um conjunto de universais ajustados previamente nos módulos que vão trabalhar apenas nos domínios que apresentarem essas propriedades. Assim, é indispensável haver uma teoria sobre os objetos a serem depreendidos dentro desse domínio de atividade, uma vez que

“a satisfação dos universais é a suposta propriedade que diferencia sentenças de outros domínios de estímulos. Quanto mais elaborada e complexa for a teoria dos universais, mais excêntrico será o domínio de estímulos para o reconhecimento da sentença. E, quanto mais excêntrico o domínio de estímulo, mais plausível é a especulação de que há a necessidade de se computar por um mecanismo especial” (Fodor, 1983: p.51)⁸⁶.

Para Fodor, os fatos relacionados à percepção da fala são um exemplo dessa excentricidade e, por conseguinte, especialidade do domínio responsável pela linguagem. Como exemplo, ele traz à luz o *Efeito McGurk* (McGurk & McDonald, 1976) no qual os resultados mostram que o reconhecimento dos fones está relacionado às intenções fonéticas do falante, fazendo com que haja uma sintonia entre os mecanismos de produção e de percepção da fala. Tais dados mostram que há “sistemas especializados para tarefas especializadas” uma vez que, por exemplo, a alucinação do *Efeito McGurk* revela que há um sistema especializado, pois a tarefa de perceber a fala exige um viés pré-determinado para se considerar algumas

⁸⁶ “Since the satisfaction of the universals is supposed to be a property that distinguishes sentences from other stimulus domains, the more elaborate and complex the theory of universals comes to be the more eccentric the stimulus domain for sentence recognition. And, as remarked above, the more eccentric a stimulus domain, the more plausible the speculation that it is computed by a special-purpose mechanism” (Fodor, 1983: p.51).

características e descartar outras, i.e., neste caso do *Efeito McGurk*, a alucinação decorre da especialização do módulo.

Uma segunda característica central para a proposta modular é sobre o encapsulamento informacional. Isso significa que o sistema de entrada não tem acesso direto a todas as informações promovidas pelos sentidos. Por exemplo, no caso da percepção visual, não há interesse pelo que se ouve. O encapsulamento teria, então, a função de confirmar as hipóteses lançadas sobre um conjunto de dados devido à falta de informações apresentadas pelos sentidos. Para Fodor, a noção de encapsulamento é importante, pois há mais coisas representadas do que disponíveis para os sistemas de confirmação.

Diante do encapsulamento dos sistemas de entrada, há a constatação de mais duas outras características como sua consequência: (i) os sistemas vão realizar operações mandatórias sobre os dados apresentados para (ii) que o sistema seja rápido.

Por fim, Fodor argumenta que os sistemas de entrada estão associados a estruturas neuronais fixas. Tais estruturas sugerem uma arquitetura natural para a ideia de encapsulamento informacional, de modo que haveria conexões estruturais que privilegiariam o acesso às informações de uma estrutura a outra que, por sua vez, garantiria a rapidez das operações modulares.

Por fim, há os processadores centrais que são a cognição que processa as informações que vêm dos sistemas de entrada. Tais informações não são as mesmas dos sentidos, trata-se de representações sobre os objetos distais obtidos pela apreciação dos objetos proximais através dos sistemas compiladores. Tais informações estão disponíveis a todo o sistema cognitivo.

Ao contrário dos sistemas de entrada, os processadores centrais não são de domínio específico, não são informacionalmente encapsulados, não apresentam estruturas neuronais fixas, são lentos e executam operações por analogia ao invés de executar operações mandatórias. O próprio Fodor argumenta que não há evidências para a modularidade dos processadores centrais. Assim, ele atribui a função de fixação de crenças a essa “parte” do seu sistema.

A título de exemplo da diferença relativa à maneira como trabalham os

sistemas de entrada e o processador central, Fodor cita, mais uma vez, o *Efeito McGurk*, pois mesmo sabendo que se trata de uma alucinação (informação esta fixada nas nossas crenças), há a manutenção do efeito, ou seja, a atividade dos sistemas de entrada não é afetada por nossas crenças. Em suas palavras:

“os sistemas de entrada parecem ser condicionados pelos estímulos primeiramente, por isso exploram processos computacionais que são relativamente insensíveis à estrutura geral do sistema de crenças do organismo. Ao passo que, quando nos voltamos à fixação de crenças, nos deparamos com problemas complexos que parecem ser intratáveis precisamente porque dizem respeito a processos mentais que não são locais” (Fodor, 1983: p.117)⁸⁷.

Em resumo, a proposta é a de que

“não há processos centrais de conteúdo específico para a performance do que correspondentemente as estruturas neurais específicas têm sido identificadas. Tudo que sabemos agora é compatível com a proposta que a solução de problema central é subservida pelos mecanismos neurais equipotenciais” (Fodor, 1983: p.119)⁸⁸,

Tendo, então, apresentado a proposta da *Modularidade da mente* de Fodor, passaremos à versão revisada da TM. Será possível verificar que a proposta de Mattingly & Liberman (1985) é influenciada pelas bases da proposta fodoriana.

2.4 A Teoria Motora Revisada

2.4.1 Proposta

A proposta da TM revisada (Liberman & Mattingly, 1985: p.237) “é motivada pelas considerações que não se aplicam necessariamente fora do domínio da fala”, i.e., tal como o propõe o paradigma gerativista, trata-se de uma teoria que vai considerar a fala como especial. E essa especialidade advém da tese modular da

⁸⁷ “[...] the input systems appear to be primarily stimulus driven, hence to exploit computational processes that are relatively insensitive to the general structure of the organism's belief system. Whereas, when we turn to be intractable precisely because they concern mental processes that aren't local” (Fodor, 1983: p.117).

⁸⁸ “To put these claims in a nutshell; there are *no* content-specific central processes for the performance of which correspondingly specific neural structures have been identified. Everything we now know is compatible with the claim that central problem-solving is subserved by equipotential neural mechanisms” (Fodor, 1983: p. 119).

mente, uma vez que são esperados “sistemas especializados para tarefas especializadas” (Fodor, 1983: p.52).

Grosso modo, os autores vão propor que, devido à falta de invariância presente nos padrões acústicos de fala sintetizada, os objetos da percepção talvez estivessem nos processos motores subjacentes ao nível acústico.

Uma pergunta a ser feita diante da proposta é a de como se perceberiam os processos motores subjacentes ao nível acústico. Em outras palavras, se a fala é veiculada em um canal acústico, como se percebem as atividades motoras subjacentes ao processo de fonação? Conforme os proponentes da TM, a resposta para essa passagem do nível acústico para o articulatório é possível com a adoção da Modularidade da Mente de Fodor (1983).

Tal como apresentado, a percepção da fala seria um sistema de entrada (o Módulo Fonético) que, através da audição (sistema compilador), receberia os estímulos proximais (as ondas acústicas) e os converteria em representações dos objetos distais para, em seguida, entregá-las ao processador central. Como se pode notar, o sistema de entrada da percepção da fala (um módulo especializado, de domínio específico, encapsulado informacionalmente, rápido e que realiza operações mandatórias) não entrega ao processador central os movimentos dos articuladores. Ele entrega uma representação dos articuladores que seria a fonte das ondas acústicas, i.e., trata-se dos comandos neuromotores pretendidos.

De acordo com os autores, este módulo fonético seria uma “estrutura neural especial, desenhada para tomar vantagem de uma relação sistemática e única entre uma exibição proximal de um órgão do sentido e alguma propriedade de um objeto distal” (p.241). A vantagem de tomar um módulo para essa função

“é que não há, primeiro, uma representação cognitiva do padrão proximal que é modalmente geral, seguido pela tradução de uma propriedade distal particular; ao invés disso, a percepção de uma propriedade distal é imediata, isso significa que o módulo fez todo o trabalho pesado” (Lieberman & Mattingly, 1985: p.241)⁸⁹.

Uma segunda pergunta a ser feita diante da proposta é por que motivo há a

⁸⁹ “A result in all cases is that there is not, first, a cognitive representations of the proximal pattern that is modality-general, followed by translation to a particular distal property; rather, perception of the distal property is immediate, which is to say that the module has done all the hard work” (Lieberman & Mattingly, 1985: p.241).

representação dos gestos pretendidos e não dos reais? Essa resposta não é encontrada facilmente dentro do texto de 1985, no entanto é possível inferi-la: trata-se de uma vinculação à proposta da Fonologia Gerativa do SPE, onde não há mais fonemas compostos por feixes de traços distintivos de base acústica. Há, por sua vez, matrizes compostas por traços de base articulatória que também não representam os movimentos dos articuladores; trata-se de traços que compõem um dado segmento e que, após as transformações necessárias no Componente Fonológico e no Fonético, vão ser implementados pelo modelo de produção da fala. Como se pode ver, ao contrário das propostas mirabolantes para contornar os problemas de codificação e decodificação encontrados no trabalho de 1967 devido à utilização do conceito de fonema e da sua definição como um feixe de traços de natureza acústica, a TM revisada deixa de lado os inúmeros estágios de regras propostos anteriormente. Assumir o gesto pretendido é assumir o traço que compõe a matriz como primitivo de análise fonológica; é aderir ao fato de que há transformações representacionais e que a forma subjacente pode ser diferente da superficial. Em suas palavras:

“essas intenções são representadas de uma maneira específica no cérebro do falante e há um módulo especializado para conduzir sem esforço o ouvinte a essa representação. De fato, o que é verdade sobre a fala a esse respeito é verdade para todas as línguas, exceto, é claro, que o objeto distal para a linguagem é alguma representação da estrutura linguística, não meramente um gesto, e que o acesso a esse objeto requer um módulo que não é meramente fonético, mas sim fonológico e sintático” (Liberman & Mattingly, 1985: p.243)⁹⁰.

Nessa citação é fácil notar a preocupação de determinar os aspectos “universais” da percepção da fala, tal como SPE pretende para a fonologia. Também é inegável o fato de que, tal como em SPE, a percepção da fala (ou todo o módulo Fonético) esteja atrelado à representação sintática e fonológica.

Diante dessa proposta, há ainda que pensar mais uma questão: já que há um

⁹⁰ “The particular claim of the motor theory is that these intensions are represented in a specific form in the talker's brain, and that there is a perceiving module specialized to lead the listener effortlessly to the representation. Indeed, what is true of speech in this respect is true for all of language, except, of course, that the more distal object for language is some representation of linguistic structure, not merely of gesture, and that access to this object requires a module that is not merely phonetic, but phonological and syntactic as well” (Liberman & Mattingly, 1985: p. 243).

módulo fonético especializado para a fala, como se resolve a competição com um módulo auditivo já que eles se utilizam dos “mesmos” estímulos captados pelo mesmo sistema compilador, a audição?

A resposta dada pelos autores é a de que há propriedades características dos estímulos proximais que só interessam a um determinado módulo, i.e., o módulo só vai atuar sobre o seu domínio específico.

Por exemplo, no caso de localização auditiva, há características particulares sobre os estímulos que vão indicar de onde vêm os sons, ou seja, a taxa particular das diferenças de tempo de chegada dos sons nas duas orelhas que vão indicar a localização da fonte sonora. É essa taxa que interessa para a localização e nada mais: “a exibição proximal e a propriedade distal que especificam apenas complementam outros aspectos que o ouvinte depreende; eles nunca competem” (Lieberman & Mattingly, 1985: p.243).

Já para a percepção fonética, embora as transições presentes no sinal da fala compartilhem características com outros sons presentes no meio ambiente, os seres humanos, em decorrência da existência dos módulos, parecem não ouvi-las como *glissandos* ou *trinados*. Tais características fazem com que os módulos fonético e auditivo estejam em competição direta, mas essa competição tende ao módulo fonético devido à especialidade do módulo fonético, i.e., ele vai tomar precedência sobre a audição uma vez que o módulo fonético vai antes de tudo tentar traçar alguma relação com os objetos distais, os gestos pretendidos; caso não consiga, o módulo auditivo já terá feito a sua análise. Esse argumento se baseia no fato de não ouvirmos *glissandos* e *trinados* durante o reconhecimento de um sinal de fala, ou seja, uma vez identificado como fala, o sinal acústico vai ser compreendido como tal. (cf. Mattingly & Liberman, 1985).

2.4.2 Evidências experimentais

Nesta seção apresentaremos as evidências experimentais que fundamentam a versão revisada da TM.

2.4.2.1 *Dados de coarticulação*

Um primeiro argumento em favor do módulo fonético para a percepção da fala vem do trabalho de Fowler (1984), no qual a autora vai tentar buscar unidades perceptuais (ou *percepts*) a partir de dados de coarticulação⁹¹. A questão é que não há uma relação entre a segmentação do sinal acústico (objeto proximal) com os *percepts* fonéticos, pois “a informação acústica para cada unidade fonética particular é tipicamente sobreposta [...] com a informação de outras unidades” (Mattingly & Liberman, 1985: p.246).

Fowler nota, então, que a percepção parece ser segmentada de uma maneira que o sinal acústico não é, pois o *percept* não espelha a sobreposição de uma informação do som. Esse fato não é explicado com a adoção de uma teoria de base auditiva, pois, como não há uma segmentação evidente do sinal acústico, como as unidades discretas emergem na percepção a partir das interações puramente acústicas?

A resposta da TM é a seguinte:

“a TM parte da assunção de que a coarticulação, e a sobreposição de informação fonética resultante no padrão acústico, é uma consequência do processo eficiente pelo qual os gestos fonéticos discretos são realizados no comportamento mais ou menos independente dos articuladores. A Teoria sugere, então, que um processo perceptual igualmente eficiente use o resultado do padrão acústico para recuperar os gestos discretos” (Mattingly & Liberman, 1985: p.246-247)⁹².

Ainda sobre coarticulação, os autores argumentam que a variação articulatória dependente de contexto é pervasiva, i.e., na representação acústica de cada categoria fonética já estudada há porções condicionadas pelo contexto que contribuem para a percepção de determinado segmento. Por exemplo, as transições sensíveis ao contexto são pistas para a identificação de oclusivas, nasais, fricativas, líquidas, semivogais e vogais. Da mesma forma, os ruídos (que também são sensíveis ao contexto) presentes em oclusivas e fricativas também são pistas para

⁹¹ É evidente que toda a fala é coarticulada. Caso contrário, a fala seria algo como o “soletramento”.

⁹² “In contrast, the motor theory begins with the assumption that coarticulation, and the resulting overlap of phonetic information in the acoustic pattern, is a consequence of the efficient processes by which discrete phonetic gestures are realized in the behavior of more or less independent articulators. The theory suggests, then, that an equally efficient perceptual process might use the resulting acoustic pattern to recover the discrete gestures” (Liberman & Mattingly, 1985: p. 246-247).

a sua identificação.

Para uma teoria de base estritamente auditiva, esse grande número de variações de pistas promoveria um sem número de assunções arbitrárias para acomodar os fatos. Já para a TM, a variação é fonte de informação que vai ser processada pelos sistemas de entrada.

Ainda pensando nessa relação entre o nível sonoro e o articulatório, os autores trazem à discussão o papel importante do silêncio na produção de diferentes *percepts*. Por exemplo, em uma sílaba como [sa], as transições entre a consoante e a vogal não são mascaradas pelo ruído fricativo (cf. Harris, 1958 *apud* Liberman & Mattingly, 1985: p.247). Esse dado mostra que, ao contrário do que sugere uma teoria auditiva, o silêncio presente nas oclusivas não tem o objetivo de proteger as transições de um possível mascaramento, já que em um contexto de fricativas e vogais não se encobrem as transições. O silêncio, em realidade, é melhor explicado em termos articulatórios:

“o silêncio age no módulo fonético especializado para informar ao ouvinte que o falante fechou completamente seu trato vocal para produzir uma consoante oclusiva, ao invés de apenas tê-lo contraído para produzir uma fricativa. Segue disso, então, que o silêncio, pela sua presença ou ausência, determinará se transições idênticas são pistas nos *percepts* que pertencem a um modo ou a outro” (Liberman & Mattingly, 1985: p.248)⁹³.

⁹³ “[...] silence acts in the specialized phonetic mode to inform the listener that the talker completely closed his vocal tract to produce a stop consonant, rather than merely constructing it to produce a fricative. It follows, then, that silence will, by its presence or absence, determine whether identical transitions are cues in percepts that belong to the one manner or the other” (Liberman & Mattingly, 1985: p.248).

2.4.2.2 Efeito *McGurk*

Um dos dados mais intrigantes (e já citado anteriormente na seção 2.3.2 como argumento para a proposta de Fodor) é o *Efeito McGurk* (McGurk & McDonald, 1976). O efeito consiste na apresentação simultânea de dois estímulos diferentes aos sujeitos do experimento. Trata-se da sincronização de uma sequência sonora como [ba, ba, ba] com a imagem de uma pessoa articulando [bɛ, vɛ, ðɛ]. O resultado perceptual é [ba, va, ða]⁹⁴.

Para a TM, o módulo fonético vai utilizar os estímulos proximais para produzir a percepção coerente de um evento distal, i.e., isso ocorre porque a percepção não é auditiva e visual, ela é gestual.

Os autores da TM argumentam que há experimentos que mostram que as crianças em fase pré-linguística preferem olhar para o rosto que está articulando o som que elas ouvem ao rosto de alguém que articula uma vogal diferente (cf. Kuhl & Meltzoff, 1982 *apud* Liberman & Mattingly, 1985: p.249) de modo que as crianças buscam vincular um dado som com a sua respectiva articulação.

2.4.2.3 *Percepção duplex*

Embora pouco mencionada, uma evidência importante (talvez a mais robusta) para a proposta da TM é a chamada percepção duplex.

O experimento é o seguinte: como auxílio de um fone de ouvido, apresenta-se a um ouvido uma sílaba (a base) sem a transição que diferencia [da] de [ga], por exemplo. No outro ouvido, é apresentada apenas essa transição. Se apresentados separadamente (um de cada vez), cada um dos estímulos é identificado com um ruído. No entanto, quando apresentados ao mesmo tempo (sincronizados), o ouvido em que foi apresentada a base percebe a sílaba completa e o outro ouve apenas um ruído, simultaneamente.

⁹⁴ É interessante notar que esta é a maneira como o *Efeito McGurk* é descrito por Liberman & Mattingly (1985: p.248). No entanto, no texto de McDonald & McGurk (1976) não há menção a dados que envolvam a produção de outras vogais e fricativas. Tal como apresentado na seção 2.3.2 do Capítulo 2 deste trabalho, seu experimento apenas apresentava os dados referentes à percepção de oclusivas bilabiais, dentais e velares.

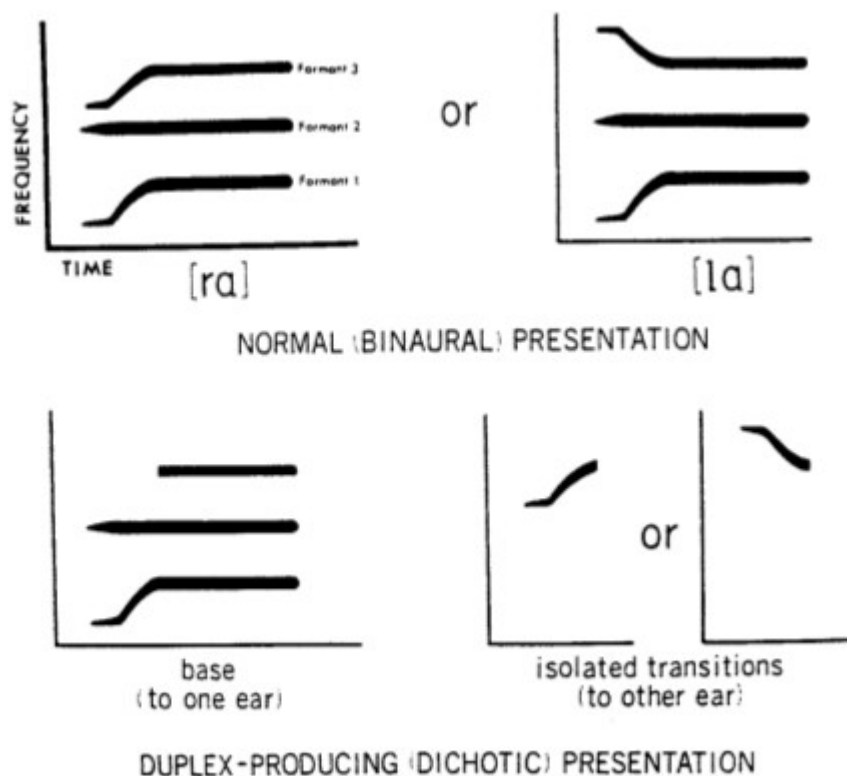


Ilustração 17: Estímulos apresentados no experimento de percepção duplex (Lieberman; Isenberg & Rakerd, 1981: p.134).

Na Ilustração 17, é possível verificar os estímulos apresentados aos informantes no experimento original de Liberman, Isenberg & Rackerd (1981). Na sua parte de cima, estão os estímulos sintetizados para a produção de [ra] e [la], respectivamente. Na parte inferior, está do lado esquerdo o estímulo apresentado sem a transição de F3 e, do lado direito, estão as transições retiradas.

Esse experimento mostra como uma mesma transição é interpretada de maneiras distintas a depender de seu contexto de produção, ou seja, ela é ora uma pista para um gesto fonético e ora um som que não é fala, um som qualquer. Trata-se de *percepts* distintos. Para os proponentes da TM, “deve haver dois módulos que podem usar de alguma maneira o mesmo *input* para produzir representações simultâneas de dois objetos distais” (Lieberman & Mattingly, 1985: p.249).

2.5 Sumário da seção

Nesta seção, apresentamos como se desenvolveu a construção da Teoria Motora da Percepção da Fala. Iniciamos o percurso pela primeira formulação da teoria (a *Early Motor Theory*), ainda na década de 1950; em seguida, passamos pela “correção no meio do caminho” de 1967, na qual os autores mudam seu olhar de um horizontalismo para um verticalismo; e, por fim, concluímos com a TM Revisada (Lieberman & Mattingly, 1985), influenciada pela proposta da Fonologia Gerativa de SPE (1968), pelos dados do feito McGurk (McDonald & McGurk, 1979) e pela Modularidade da Mente de Fodor (1983).

Na próxima seção (3 do Capítulo 2), apresentaremos a proposta da Teoria do Realismo Direto da Percepção da Fala e tentaremos estabelecer uma relação com o advento das abordagens fonológicas dinâmicas, como a Fonologia Articulatória⁹⁵ de Browman & Goldstein (1992).

⁹⁵ Em outros trabalhos (eg. Nishida, 2009), decidimos utilizar o termo Fonologia Gestual ao invés de Fonologia Articulatória. Naquele momento, havia a necessidade de se salientar o aspecto gestual e fonológico da proposta que, em leituras apressadas, poderiam ser consideradas como uma Fonética Articulatória. Aqui, decidimos manter o termo Fonologia Articulatória como um indício histórico da sua formulação, i.e., nesse período do início da sua formulação, a teoria tinha esse nome: Fonologia Articulatória (*Articulatory Phonology*).

3 A PERCEPÇÃO DA FALA EM UM ARCABOUÇO DINÂMICO

“Não adianta fingir que não sente

Gente sente tudo!

Gente tem que se envolver

A gente é emoção.

Corremos com as pernas e vamos devagar com o coração“

(Alexandre Nero. Cadê meu Jardim?)

Nesta seção, iremos apresentar o percurso da Teoria do Realismo Direto da Percepção da Fala (doravante TRD). Partiremos dos seus primeiros achados sobre a importância de aspectos coarticulatórios para a percepção da fala. Em seguida, traremos à luz a teoria de percepção que embasa a proposta (Gibson, 1966) para, depois, resenhar a sua primeira formulação (1986).

No decorrer desta seção, ainda passaremos pelos conceitos-chave de *gesto articulatório* da Fonologia Articulatória⁹⁶ (Browman & Goldstein, 1992) e de *paridade* (Liberman & Whalen, 2000) para chegarmos à proposta de Goldstein & Fowler (2003) que unifica o tratamento da produção e percepção da fala em um arcabouço dinâmico.

3.1 Os primeiros achados

No início da década de 1980, Fowler (1981) publica um trabalho no qual são discutidas questões sobre a percepção a partir de dados reais e coarticulados. A tentativa de se deixar de lado os experimentos conduzidos com o auxílio de sintetizadores para a criação de estímulos sonoros revela uma busca por outros parâmetros importantes para a percepção da fala, uma vez que os experimentos de

⁹⁶ Neste trabalho, escolhemos chamar a proposta de Browman & Goldstein (1992) de Fonologia Articulatória ao invés de Fonologia Gestual. A nossa escolha é muito mais histórica do que teórica, i.e., decidimos manter a tradução “direta” do nome do modelo com o intuito de recuperar as origens da teoria que, de acordo com a nossa proposta, são fundamentais para a constituição da proposta de base articulatória do Realismo Direto de Fowler (1986 e seguintes). Cabe mencionar, contudo, que chamá-la de Fonologia Gestual evitaria alguns equívocos. Por exemplo, Peter Ladefoged (1990), ao comentar o texto de Browman & Goldstein (1990), não compreendeu muito bem qual era a proposta da FAR, pois confundiu o que seria uma *fonologia articulatória* com uma *fonética articulatória*.

Liberman e colegas (cf. Liberman, 1957) já haviam mostrado a insuficiência do tratamento puramente auditivo sobre os dados referentes ao ponto de articulação de consoantes oclusivas.

Nesse trabalho, a autora⁹⁷ rompe com a busca pela invariância dos segmentos que desconsiderava a coarticulação⁹⁸ como constitutiva da produção da fala, i.e., ela elabora um experimento que examina a influência que a coarticulação exerce sobre as vogais e as consoantes. A sua preocupação é a de mostrar que os padrões coarticulatórios presentes na produção dos segmentos são indicadores sobre as estratégias articulatórias planejadas e controladas pelos falantes. Para tanto, Fowler examina as relações coarticulatórias entre vogais não adjacentes que influenciam as consoantes.

Ao conduzir um experimento sobre a produção, a autora verifica que há efeitos coarticulatórios à distância da vogal acentuada sobre a vogal átona precedente ou seguinte⁹⁹. Ela ainda relata que a vogal acentuada que precede uma consoante exerce influência mais forte que as vogais que seguem as acentuadas. Tais resultados foram obtidos a partir da medição de F2 das vogais em sílabas adjacentes.

Para se chegar a esse resultado sobre a produção das vogais de sílabas acentuadas e não acentuadas de palavras do inglês, a autora considera que a coarticulação entre segmentos adjacentes e o efeito chamado “encurtamento compensatório” são em realidade um mesmo efeito da coprodução de vogais e consoantes. Essa proposta advém da literatura fonética sobre a produção de vogais e consoantes que diz que as vogais exercem influência sobre a parte do sinal acústico que normalmente é atribuído à realização de consoantes e que, da mesma forma, as consoantes também devem ser consideradas como sendo sobrepostas,

⁹⁷ Neste momento do texto, tenho tratado Fowler (1981 e 1984, nas próximas páginas do trabalho) como a autora, pois a Teoria do Realismo Direto da Percepção da Fala somente é proposta em 1986. Até esta data, Fowler apenas sugeria que era necessário considerar um tratamento realista para a produção da fala.

⁹⁸ Para Fowler, coprodução é um fenômeno distinto de coarticulação. A coarticulação está relacionada aos efeitos “automáticos” da produção de um dado som próximo a outro. Por sua vez, a coprodução está relacionada à própria realização coordenada dos articuladores envolvidos na produção da fala (Fowler, 1980).

⁹⁹ Resultados semelhantes a esse já haviam sido documentados por Öhman (1966). Nesse trabalho, o autor apresenta evidências experimentais que apontam para a produção simultânea do *onset* de uma vogal (V1) e de uma consoante (C) em sequências V1CV2.

parcial ou totalmente, às vogais adjacentes¹⁰⁰ (cf. Carney & Moll, 1971; Kent & Moll, 1972; Ohman, 1966; Perkel, 1969; *apud* Fowler, 1981: p.128). Nas palavras da autora:

“a coprodução de uma vogal com uma consoante conduzirá a medida de encurtamento da vogal com relação à duração de uma vogal quando produzida isoladamente. Por hipótese, a vogal é medida como mais curta não necessariamente porque ela é mais curta em qualquer sentido articulatorio, mas porque a maior extensão de duração sobre a qual ela é coproduzida com a consoante é convencionalmente atribuída à consoante apenas quando as medidas são realizadas. Nessa perspectiva, o encurtamento compensatório é bidirecional simplesmente porque as influências coarticulatórias de vogais sobre consoantes se estendem nas duas direções” (Fowler, 1981: p.128)¹⁰¹.

Para ilustrar essa relação bidirecional entre vogais e consoantes, a autora lança mão de uma ilustração para mostrar que o encurtamento das vogais é aparente devido à sobreposição de uma consoante, estando ela no ataque ou na coda da sílaba. Nessa ilustração 18, é possível verificar que o encurtamento da vogal é aparente. Ao se olhar para a articulação, a vogal permanece do mesmo “tamanho”. O que de fato ocorre é uma sobreposição da consoante sobre o espaço acústico e articulatorio da vogal.

¹⁰⁰ A noção de que a cadeia da fala é organizada pela produção de vogais nas quais as consoantes se sobrepõem é importante para que ocorra uma mudança de olhar sobre a cadeia da fala. Esse ponto será adotado principalmente nas primeiras versões da Fonologia Articulatória. Nas pautas gestuais apresentadas naqueles textos, os sons vocálicos organizavam a cadeia da fala de modo que os todos os outros gestos estavam “submissos” ao de dorso. Para um histórico sobre o desenvolvimento da FAR, ver Silva (2002).

¹⁰¹ “[...] coproduction of a vowel with a consonant will lead to a measured shortening of the vowel relative to the vowel's duration when produced in isolation. By hypothesis, the vowel is measured to be shorter not necessarily because it is shorter in any articulatory sense, but because most of the durational extent over which it is coproduced with a consonant is conventionally assigned only to the consonant when measurements are made. On this view, compensatory shortening is bidirectional simply because coarticulatory influences of vowels on consonants extend in both directions” (Fowler, 1981: p. 128).

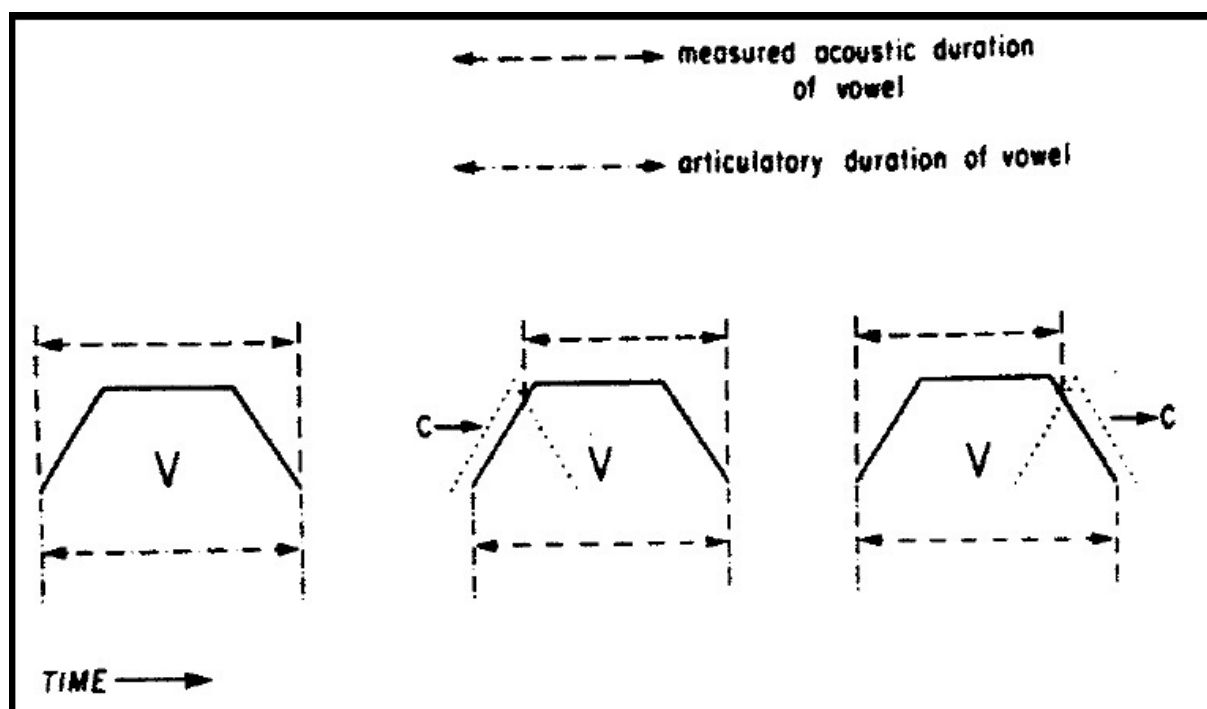


Ilustração 18: Representação esquemática da coprodução de vogais e consoantes (Fowler, 1981: p.46).

O experimento conduzido derivou da observação de que os efeitos de encurtamento relacionados às vogais acentuadas são organizados em camadas, i.e., vogais acentuadas são encurtadas não apenas pela consoante precedente ou seguinte, mas também pelo contexto de sílabas não acentuadas.

Desta forma, para que coarticulação e encurtamento sejam considerados como coprodução, a previsão do experimento é a de que vogais não acentuadas devem ser afetadas acusticamente quando realizadas em contexto vocálico acentuado, de modo que os efeitos de coarticulação perseveratória sejam mais fortes que os antecipatórios caso a vogal acentuada preceda a não acentuada¹⁰².

Esses achados sobre os efeitos coarticulatórios entre vogais fazem com que

¹⁰² Conforme Ladefoged & Johnson (2011: p. 70), a coarticulação antecipatória é caracterizada pela antecipação do movimento de um articulador que será utilizado na produção de um som seguinte. Para tanto, eles citam como exemplo a antecipação da aproximante /w/ quando é o segundo elemento dos encontros consonantais como *twice* ou *quick*. Nesses exemplos, é possível identificar a protrusão labial ocorrendo durante a oclusão de /t/ e de /k/. Já a coarticulação perseveratória (que é ocorre bem menos que a antecipatória em inglês – cf. Ladefoged, 2011: p. 92) é caracterizada pela manutenção de uma articulação sobre o segmento seguinte. Por exemplo, “it is” (que pode ser pronunciado [it.iz]) é frequentemente pronunciado como [its], como resultado da coarticulação perseveratória da oclusiva surda /t/.

se inicie uma discussão sobre quais estratégias (se há alguma) perceptuais são utilizadas para a identificação de vogais coarticuladas.

A ideia do experimento de percepção de vogais coarticuladas tem por objetivo verificar se há contextos que impedem ou que promovem a percepção de uma vogal, dado que (conforme o experimento de produção) uma vogal acaba sendo influenciada por outra à distância e, devido a isso, poderia haver alguma “confusão” quanto à sua qualidade. É claro que essa questão só é relevante se se considerar que os falantes utilizam informações coarticulatórias como pistas para a percepção da fala. E este é o olhar inovador que Fowler lança sobre a percepção da fala.

O experimento consistia em testar a percepção sobre vogais não acentuadas com relação à vogal acentuada da sílaba precedente, i.e., tal como foi encontrado no experimento de produção, a experimentadora esperava que os ouvintes utilizassem pistas da coarticulação da vogal acentuada sobre a não acentuada nos julgamentos relativos à qualidade da vogal. Por exemplo, ao “trocar” os contextos adjacentes às vogais não acentuadas, há um aumento nos julgamentos de similaridade quando o contexto vocálico é o mesmo. Por exemplo, na sílaba medial do trissílabo ['i_i.bΛ_i.bi], a vogal é considerada a mesma se as ocorrências vocálicas forem ouvidas em sequência devido ao fato de o som ter sido produzido entre [i], ou seja, trata-se de um [Λ] produzido diante de [i] e as outras sílabas também obedecem ao mesmo contexto¹⁰³. Em uma sequência do tipo ['i_i.bΛ_i.bi] e ['i_a.bΛ_u.bi], as sílabas mediais são tidas como diferentes pelo fato de [Λ] ter sido produzido diante de [i] em um caso e diante de [u] em outro. Há também que se considerar que a vogal precedente ([i]) foi produzida diante de [i] no primeiro exemplo e diante de [a] no outro.

Tais resultados mostram que as vogais não acentuadas são consideradas dentro do domínio das vogais acentuadas e a sua percepção é “facilitada” pelo contexto e não pelas características isoladas (ou inerentes) de um dado segmento.

Ainda sem explicar muito bem como uma abordagem realista poderia ajudar a entender os fatos encontrados, Fowler finaliza o texto argumentando que uma

¹⁰³ O [i] subscrito se refere ao contexto em que a sílaba foi originalmente produzida. Por exemplo, [i_i] significa que trata-se de um [i] produzido diante de outro [i] e [a_i], uma [a] diante de [i]. Tais estímulos foram produzidos a partir da “colagem” de produções realizadas anteriormente, uma vez que seria impossível realizar um [Λ] com as características diante de um [u], sendo que ele está entre [i].

maneira possível de interpretar esses achados é a proposta de Gibson (1966), na qual

“as propriedades sistemáticas constituem o estímulo da percepção; na medida em que o percebedor registra a estrutura de um estímulo acústico, a estrutura que corresponde aos gestos articulatórios é registrada, e as categorias fonéticas percebidas tenderão a uma definição articulatória” (Fowler, 1981: p.138)¹⁰⁴.

Em um trabalho seguinte (Fowler, 1984), a autora encontra alguns fatos que apontam mais uma vez para uma abordagem realista sobre a percepção da fala. Nesse trabalho, a autora mostra que há estudos que argumentam que à medida que os falantes de uma dada língua se tornam usuários proficientes, eles deixam de considerar toda a informação fonética presente no sinal acústico da fala, pois os usuários reconhecem sílabas e palavras mais rapidamente do que segmentos fonéticos (cf. McNeil & Liding, 1973; Savin & Bever, 1970 *apud* Fowler, 1984: p. 359). No entanto, ela argumenta que a estrutura fonética das palavras é psicologicamente real, porque tanto nos casos de erros de fala como nos jogos de linguagem o que acontece são desordenamentos e substituições dos segmentos ou operações fonéticas (fonológicas) sobre os segmentos que estruturam as palavras. Como decorrência dessa observação, a sua hipótese é a de que o sinal acústico da fala deve promover informação suficiente para que se extraia a estrutura fonética da mensagem pretendida pelo falante.

Embora essa hipótese seja tida como logicamente possível (e até mesmo “óbvia” para os menos avisados), há problemas em se considerar o sinal acústico como o provedor de informações necessárias e suficientes para a percepção de uma mensagem. Segundo a autora, as barreiras para essa abordagem são os problemas da segmentação e da invariância, “pois os gestos articulatórios para os sucessivos segmentos fonéticos não são temporalmente discretos” (Fowler, 1984: p.359) devido à fala ser coarticulada.

O problema da segmentação se refere à dificuldade de separar a cadeia da

¹⁰⁴ “The systematic properties of the source – its structure or its changing structure over time – are reflected in the systematic properties of the proximal (acoustic) stimulus. These systematic properties constitute the stimulus for perception; to the extent that a perceiver registers the structure in the acoustic stimulus, structure that corresponds to articulatory gestures is registered, and perceived phonetic categories will tend to have an articulatory definition” (Fowler, 1981: p.138).

fala em segmentos do tamanho de um fone, sendo que as informações relativas a tais segmentos se sobrepõem no tempo. Por sua vez, o problema da invariância está relacionado às tentativas de descobrir quais são os invariantes acústicos que existem entre os *tokens* que fazem os falantes os reconhecerem como um mesmo *type*. A dificuldade reside justamente no fato de que tais informações não são “estáveis”, são sensíveis ao contexto. Diante de tais problemas, a autora decide, então, testar dois modos possíveis de segmentar o sinal acústico da fala, pois essas perspectivas oferecem explicações diferentes sobre como ocorre a segmentação da fala.

Na Ilustração 19, é possível visualizar as hipóteses de segmentação que são testadas no experimento de Fowler. Em (a), há uma esquematização de como a cadeia da fala é produzida. Cada arco representa o início e o término da execução de um segmento. O tempo é representado no eixo horizontal, enquanto a proeminência é descrita no vertical. Segundo a autora, entende-se por proeminência a parte de um segmento em que as suas características acústicas estão mais salientes no sinal. Por exemplo, em uma sílaba como /si/, o /s/ é mais proeminente que /i/ durante o ruído fricativo. Na ilustração, é possível esquematizar essa proposta ao se propor que mesmo se o /i/ começar a ser produzido durante a constrição da fricativa (representado pela parte hachurada), a sua proeminência vai ser menor que a de /s/.

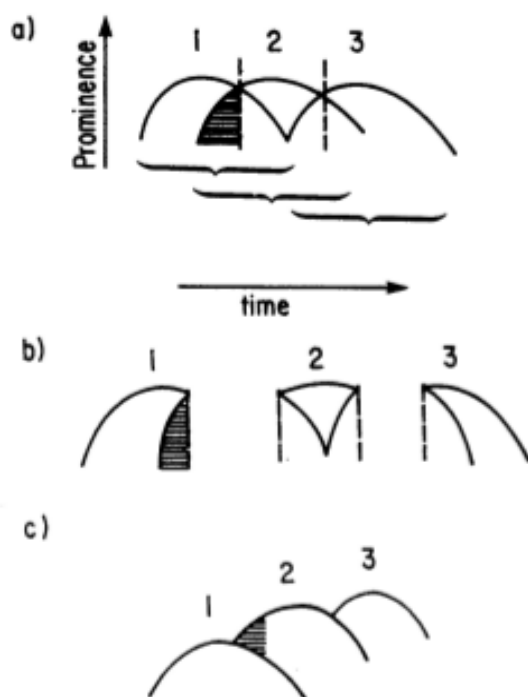


Ilustração 19: Esquemas de segmentação do sinal da fala (Fowler, 1984: p.360).

Em (b), temos uma primeira estratégia de segmentação do sinal acústico. Nesta, leva-se em conta “a proeminência relativa dos segmentos sucessivos para se estabelecer as fronteiras entre eles” (Fowler, 1984: p.360), i.e., este procedimento divide o sinal acústico em segmentos fonéticos sensíveis ao contexto, pois é possível verificar que há informações de menor proeminência dos segmentos durante a produção do mais proeminente.

Já em (c), temos a segunda maneira de segmentar a fala. Trata-se de uma segmentação que leva em conta a coarticulação (a sobreposição) dos segmentos fonéticos. Ao contrário de (b), essa hipótese considera que os segmentos são livres da influência dos contextos fonéticos vizinhos. Por exemplo, o /i/ de uma sílaba como /si/ é identificado a partir do início da sua produção e não quando ele atinge o seu ponto de maior proeminência.

A autora relata que a maneira de segmentação (b) é a mais adotada na literatura, mas isso não significa que seja assim que os falantes segmentam o sinal acústico. Trata-se sobretudo de uma estratégia de aumentar a confiança sobre as medidas realizadas sobre os segmentos. Diante disso, desenvolveram-se duas

linhas de pesquisa que evidenciam o fato de que os usuários da língua não segmentam o sinal dessa maneira.

A primeira linha de pesquisa é aquela que aponta para os achados de que os ouvintes usam a coarticulação antecipatória de um segmento sobre outro como fonte de informações para a identificação do segmento influenciado. Por exemplo, Whalen (1984 *apud* Fowler, 1984: p.360) recortou e trocou os ruídos fricativos das sílabas /sa/, /su/, /ja/ e /ju/ e, ao solicitar que os informantes identificassem as vogais, verificou que os ouvintes foram mais rápidos e precisos quando os ruídos forneceram informações antecipatórias sobre as vogais do que quando havia desencontro de informações. Em outras palavras, ao se assumir a segmentação sensível ao contexto de (b), é possível formular a hipótese de que os ouvintes utilizam pistas que precedem o alvo (a proeminência) de um segmento para prever a sua identidade.

Uma segunda linha de pesquisa mostra que os ouvintes realizam compensações quando se identificam influências coarticulatórias. Por exemplo, um dos aspectos acústicos que diferenciam as fricativas /s/ e /ʃ/ é a localização da energia de produção no espectro, sendo mais alta para /s/ que para /ʃ/. No entanto, ambas as fricativas têm sua energia abaixada diante de /u/ devido à antecipação do arredondamento dos lábios. Tais fatos sobre a produção dessas fricativas são compatíveis com o fato de que um mesmo ruído fricativo ora é percebido como /s/ ora como /ʃ/ se ele for seguido de uma vogal arredondada, i.e., um mesmo ruído fricativo é identificado como /ʃ/ diante de /i/, mas como /s/ antes de /u/ (cf. Mann & Repp, 1980 *apud* Fowler, 1984: p.361). Tal como ilustrado em 19, esse fato é explicado a partir da esquematização em (b), uma vez que se assume uma segmentação discreta e sensível ao contexto. Em outras palavras, a compensação reflete os ajustes que são realizados a partir do conhecimento que se tem sobre os efeitos de antecipação de um segmento sobre o outro.

Contudo, Fowler salienta que, se considerarmos juntos os dois fatos que embasam a proposta de segmentação (b), temos uma situação paradoxal. No caso de uso da informação antecipatória, suporíamos que os efeitos coarticulatórios de um segmento deveriam ser identificados como tais antes do seu *onset*. Caso contrário, o tempo de reação de identificação do segmento antecipado não seria

reduzido quando a informação coarticulatória for preditiva. Mas, nesses experimentos, os informantes são em realidade mais rápidos. Já no caso dos efeitos de compensação, suporíamos que um segmento que segue um alofone sensível ao contexto seria usado para guiar a identificação da influência contextual do alofone. Nas palavras da autora:

“[...] , em uma instância, a informação coarticulatória facilita a identificação *tardia* do segmento que o antecipa; em outra, ela [a informação coarticulatória] pode ser identificada como uma influência coarticulatória de um alofone apenas *depois* que tiver sido identificado o segmento que a antecipa” (Fowler, 1984: p.361)¹⁰⁵.

Em contrapartida, a proposta de segmentação em (c) satisfaz os dois conjuntos de achados apresentados anteriormente, pois os dados apontam para o fato de que os ouvintes segmentam a cadeia da fala considerando as suas características coarticulatórias presentes na sobreposição dos segmentos fonéticos. Como consequência dessa estratégia de segmentação, fazemos a previsão de que a informação sobre a coarticulação antecipatória é percebida como o *onset* do segmento que a antecipa e, por isso, a mesma informação não é integrada com a informação simultânea do segmento precedente. Em outras palavras, o efeito de compensação ocorre como um subproduto da estratégia de segmentação, pois a informação do segmento seguinte já está disponível em seu *onset* e não compete com a informação do segmento em execução, i.e.,

“devido às fontes de sensibilidade ao contexto não estarem integradas com a informação dos segmentos com os quais elas coocorrem no tempo, é previsto que o mesmo segmento fonético em contextos coarticulatórios distintos seja percebido aproximadamente da mesma forma pelos ouvintes. Isto é, os ouvintes devem perceber os *tokens* de um *type* fonético como iguais ou muito parecidos através dos distintos contextos fonéticos devido às fontes de influência contextual não terem sido integradas aos *tokens*” (Fowler, 1984: p.361)¹⁰⁶.

Levando em conta essas questões, Fowler elabora um experimento que

¹⁰⁵ “That is, in the one instance, the coarticulatory information facilitates *later* identification of the segment it anticipates; in the other, it can be identified as a coarticulatory influence on an allophone only *after* the segment it anticipates has itself been identified” (Fowler, 1984: p.361).

¹⁰⁶ “Because sources of context sensitivity are not integrated with information for segments with which they co-occur in time, the same phonetic segment in different coarticulatory contexts is predicted to sound approximately the same to listeners. That is, listeners may perceive the tokens of a phonetic type as the same or very similar across different phonetic contexts because sources of contextual influence have not been integrated with the tokens” (Fowler, 1984: p. 361).

pretende verificar (i) se os ouvintes usam a informação coarticulatória de uma vogal sobre o domínio acústico de um segmento que a precede e (ii) se há evidência para os efeitos de compensação. Diante disso, se os ouvintes exibirem o mesmo comportamento para as mesmas sílabas, teremos uma evidência para a segmentação em (b), de modo que os ouvintes estariam utilizando as informações coarticulatórias antes mesmo da identificação do segmento que a informação coarticulatória antecipa, mesmo que seja exigido o ordenamento inverso para explicar a compensação. Por sua vez, a estratégia (c) oferece tratamento unificado para os achados.

Em seu experimento, a autora utiliza como material a gravação das sílabas [gi] e [gu]. A partir dessas gravações, são produzidos dois tipos de estímulos para serem utilizados nos testes (que serão descritos a seguir). Primeiramente, utilizaram-se somente o *burst* e uma porção vocálica da vogal de cada sílaba, separadamente. O segundo conjunto de estímulos era composto de sílabas completas: as originais; as intra-cortadas (i.e., trocou-se o *burst* de uma sílaba pelo *burst* de uma sílaba do mesmo tipo. Por exemplo, em uma sílaba [gi], trocou-se o seu *burst* pelo de outra sílaba [gi]. Esta estratégia gerou uma sílaba [gi].); e as inter-cortadas (i.e., trocou-se o *burst* de uma sílaba pelo *burst* de uma de outro tipo. Esta estratégia gerou uma sílaba [gu], ou seja, trata-se de [g] de [gi] e um [u] de [gu].). Passemos aos experimentos.

No experimento de identificação, os ouvintes conseguiram identificar (com probabilidade acima da chance) as consoantes a partir dos *burts* isolados. Da mesma forma, eles conseguiram identificar acuradamente a vogal a partir das porções vocálicas. A autora também relata que os sujeitos do experimento também conseguiram identificar a vogal a partir do *burst* isolado com probabilidade que supera a chance.

Com relação às sílabas completas, os sujeitos realizaram a identificação correta nos estímulos originais e intra-cortados. Já nos estímulos inter-cortados, houve casos em que os ouvintes identificaram [gu] como [b], tal como relatam Liberman e colegas (1952), e como [d]. A autora também apresenta o fato de que alguns informantes relataram a produção de ditongos na sílaba [g_{ui}].

Nesse teste, os dados mostram que a informação contida no *burst* e na vogal é interpretada como integrada quando é apresentada em sílabas completas e como diferente quando apresentada isoladamente. Os dados também mostram que os ouvintes são sensíveis à informação vocálica durante a soltura da consoante.

Com relação ao tempo de reação, é possível notar que os informantes foram significativamente mais lentos na identificação das sílabas inter-cortadas. Ainda mais, ao fazer uma relação com a qualidade vocálica, os informantes se mostraram mais lentos com as sílabas do tipo [g_ui] do que com as [g_iu]. Esse resultado condiz com o fato de alguns sujeitos terem ouvido ditongos na produção de [g_ui]. Por sua vez, isso indica que o tempo de reação é diminuído porque o sujeito tem que escolher a produção ouvida como monotongo, mesmo tendo ouvido um ditongo¹⁰⁷. Esse fato mostra que os sujeitos que ouvem os ditongos percebem a informação vocálica no *burst*. Da mesma forma, aqueles que ouviram a produção como um monotongo tiveram seu tempo de reação aumentado. Isso aponta para o fato de que esses sujeitos também ouviram a informação vocálica no *burst* (caso contrário, não haveria aumento no tempo de reação), mas a informação da vogal se sobrepõe a presente no *burst*.

Por fim, a autora conduziu um teste em que os sujeitos iriam julgar pares de sílabas como iguais ou diferentes. Soma-se a essa atividade um julgamento de confiança, que tem por objetivo traçar um grau de “certeza” sobre a resposta dada pelo informante.

Dois tipos de estímulos foram utilizados. Nos do tipo A, eram apresentados dois pares de sílabas e em um dos pares haveria uma sílaba inter-cortada. Por exemplo, ouvir-se-iam os pares [g_ii]/[g_ii] e [g_ii]/[g_ui] para verificar se os sujeitos ouvem os *bursts* da segunda dupla como diferentes. Nos do tipo B, a única diferença é que o contexto vocálico não permanece o mesmo. Por exemplo, apresentaram-se aos informantes os pares [g_ii]/[g_iu] e [g_ii]/[g_uu]. Esse teste tem por objetivo verificar a estratégia de segmentação utilizada na percepção da fala, i.e., se os participantes do teste escolherem o primeiro par (*bursts* idênticos) como mais parecidos que o segundo (*bursts* diferentes), isso indica que se utiliza a informação antecipatória

¹⁰⁷ No experimento, havia três grupos de sujeitos. Em apenas um dos grupos havia a possibilidade de se considerar as produções como ditongos.

presente no *burst* para se perceber a vogal, sendo um argumento para a estratégia de segmentação em (b).

Com relação aos estímulos do tipo A, os informantes ouviram como iguais as sílabas que possuíam o *burst* igual. Esse resultado mostra que a informação da coarticulação antecipatória é audível nos contextos silábicos em questão. Em contrapartida, nos estímulos do tipo B, os informantes escolheram como mais semelhantes os estímulos com *bursts* diferentes. Quando se leva em conta os julgamentos de confiança, tais resultados se mostram “espelhados”: para os estímulos A, os informantes têm mais confiança nos seus julgamentos com relação aos *bursts* serem considerados acusticamente iguais; já para B, os sujeitos têm menos confiança sobre os pares que eles consideraram como acusticamente iguais.

Tais resultados podem ser sumarizados da seguinte maneira: os ouvintes percebem as informações coarticuladas presentes no sinal acústico, pois ouvem *bursts* distintos e conseguem identificar qual é a vogal a partir de um *burst* reproduzido isoladamente. Com relação à estratégia de segmentação, os resultados sugerem a estratégia (c), pois os resultados de tempo de reação e os julgamentos de confiabilidade sugerem que as informações sobre os segmentos estão disponíveis a partir do momento em que eles começam a ser produzidos. A hipótese (b) de segmentação sugere que apenas se percebe o de maior proeminência. No entanto, isso não ocorre.

Um ponto interessante sobre esses dados se refere à negação do fato de que Hockett (1955) comparou os efeitos de coarticulação sobre os segmentos fonéticos ao envio de uma linha de produção de ovos de Páscoa a um espremedor, i.e., a coarticulação é vista como um processo que destrói a “coerência individual dos segmentos fonéticos e a sua separação de um com outro” (Fowler, 1986: p.366). Tais resultados mostram justamente o contrário: há a percepção de informações coarticuladas. Fowler ainda cita a proposta de Liberman & Studdert-Kennedy (1978) que sugere que, como o significado das sentenças é extraído a partir de grupos coerentes de palavras (ao invés de palavra por palavra), a coarticulação é necessária para que a fala seja transmitida rapidamente. Um segundo ponto reside no fato de que a quantidade de segmentos não pode exceder uma taxa em que os ouvintes consigam recuperar o ordenamento dos segmentos produzidos e a

coarticulação teria justamente a função de permitir que segmentos de maior duração consigam ocupar pequenos intervalos de tempo. Esse aspecto, é claro, só é relevante para ouvintes que conseguem perceber as sobreposições de segmentos como sobreposições, ao invés de segmentos discretos e sensíveis ao contexto.

Como se pode perceber, tanto os achados de 1981 como os de 1984, convergem para um mesmo ponto: os ouvintes percebem informações coarticuladas, i.e., os ouvintes percebem e segmentam o sinal de maneira real, pois recuperam os segmentos produzidos pelos falantes. Em outras palavras, há a superação de uma busca pela “essência” de estruturas estáticas, de modo que se passa a considerar a percepção de aspectos dinâmicos. Esses resultados apontam para um tratamento realista sobre a percepção da fala e ele ganha maior “fôlego” quando se considera uma teoria de percepção realista para a percepção. Diante disso, apresentaremos na seção seguinte (3.2 do Capítulo 2) a base da TRD (Gibson, 1966) para, em seguida, passarmos à sua primeira formulação de 1986 (seção 3.3. do Capítulo 2).

3.2 A base da TRD: Os sentidos como sistemas perceptuais (Gibson, 1966)

A obra “Os sentidos como sistemas perceptuais” (doravante *Os sentidos*) de James Gibson (1966) inaugura uma reflexão de base ecológica acerca da maneira como os animais (humanos ou não) percebem o ambiente em que vivem. Por mais que sua proposta não verse especificamente sobre a percepção da fala, a sua abordagem promove um novo olhar sobre como ela funciona, tanto que os trabalhos da TRD vão se valer dessas considerações para a sua formulação teórica. O *pulo do gato* da proposta gibsoniana é apontar para uma interpretação multimodal acerca da percepção, i.e., para se perceber o ambiente, os seres se valem dos sistemas que estiverem disponíveis em determinada situação de sobrevivência. Esse ponto será de suma importância para que a TRD consiga imprimir um olhar sobre a percepção da fala que abandona a noção de especialidade, tão cara à TM.

Neste pequeno resumo sobre seu trabalho, apresentaremos: 1) o que há de novo no tratamento sobre os sentidos; 2) a maneira como os seres obtêm estímulos

do ambiente; 3) o que são os sistemas perceptuais; e 4) um sumário da proposta da teoria de coleta de informações.

3.2.1 O que são os sentidos?

Gibson (1966) comenta que há pelo menos duas maneiras de definir o que significa *sentir*. Podemos concebê-lo ou como sinônimo de *detectar algo* ou como *ter uma sensação* sobre algo¹⁰⁸. A proposta disponível em *Os sentidos* é a de que *sentir* significa *detectar*, i.e., os sentidos obtêm informações sobre o ambiente sem a intervenção de um processo intelectual que proporcione sensações conscientes aos seres. Em outras palavras, sua abordagem vê a percepção sem sensação (*[sensationless perception]*), mas isso não significa que ela ocorre sem a presença de estímulos. Isso somente significa que não há percepção extrassensorial, ou seja, “pode haver percepção sem sensação, mas não percepção sem informação” (Gibson, 1966: p.2).

Há toda uma tradição na psicologia, fisiologia e filosofia de realizar experimentos considerando os sentidos como receptores passivos que acarretam sensações específicas nos seres. No entanto,

“a questão sem resposta da percepção dos sentidos é como um observador, animal ou humano, pode obter percepções constantes do cotidiano tendo como base a mudança de sensações continuamente. O fato é que os animais e os homens percebem e respondem às propriedades permanentes do ambiente, assim como para as alterações que ocorrem nele” (Gibson, 1966: p.3)¹⁰⁹.

Por exemplo, uma mesa é constantemente percebida como tal mesmo havendo mudança de tamanho, iluminação, perspectivas e materiais. Para Gibson, há invariantes nesses objetos (e, conseqüentemente, no ambiente) que fazem os seres perceberem-nos como tais. Pouco importa, em sua abordagem, se uma mesa é vermelha ou laranja, se é de plástico com emblemas de marcas de cerveja ou de

¹⁰⁸ Gibson define sensação, a partir de Reid (1975 *apud* Gibson 1966: p.1), como conceitos e crenças que obtemos conscientemente sobre os nossos sentidos.

¹⁰⁹ “The unanswered question of sense perception is how an observer, animal or human, can obtain constant perceptions in everyday life on the basis of these continually changing sensations. For the fact is that animals and men do perceive and respond to the permanent properties of the environment as well as to the changes in it” (Gibson, 1966: p.3).

madeira de lei sem nenhuma inscrição. O que importa é que os seres a percebem como uma mesa independentemente das alterações realizadas nelas e das sensações que tais materiais provocam neles. Em outras palavras, a percepção constante dos objetos depende de uma habilidade de detectar as invariantes e de não se importar com a mudança de sensações. Isso só é possível ao se considerar os sentidos como detectores de informações.

Ainda, em sua proposta de que os sentidos detectam informações dos objetos no ambiente, Gibson critica a teoria largamente aceita de que os nervos que se ligam aos órgãos dos sentidos levam ao cérebro as sensações percebidas. Ele não discute que há a transmissão de informações através dos nervos. A questão fundamental para sua teoria é a de que essa noção promove um olhar passivo sobre os órgãos, pois eles fariam parte dos organismos e receberiam passivamente o que lhes for apresentado pelo ambiente. Gibson sugere que os órgãos do sentido são receptores sensoriais de natureza exploratória que possuem a capacidade de ressoar as informações externas interceptadas e, por consequência, detectadas no ambiente. Não se trata de uma transmissão de dados que se constrói em nosso cérebro, i.e., o *input* dos estímulos nos nervos sensoriais não são a base da percepção, mas apenas a base para impressões passivas dos sentidos.

Levando em conta a noção aqui colocada sobre os receptores dos sentidos como sendo ativos e acerca das relações estabelecidas entre os nervos e o cérebro, Gibson sumariza sua proposta da seguinte forma:

“Os sentidos ativos não podem ser simplesmente os iniciadores dos *sinais* nas fibras nervosas ou *mensagens* no cérebro; ao invés disso eles são análogos aos tentáculos e às antenas. E a função do cérebro quando atingido pelos seus órgãos perceptuais não é a de decodificar sinais, nem de interpretar mensagens, tampouco aceitar imagens. Essas velhas analogias não se aplicam mais. A função do cérebro não é também *organizar* o input sensorial ou *processar* os dados, na terminologia moderna. Os sistemas perceptuais, incluindo os centros nervosos em seus vários níveis no cérebro, são maneiras de procurar e extrair informação sobre o nicho a partir da matriz de fluxo da energia do ambiente” (Gibson, 1966: p.5 – *grifos* do original)¹¹⁰.

¹¹⁰ “The active senses cannot be simply the initiators of *signals* in nerve fibers or *messages* to the brain; instead they are analogous to tentacles and feelers. And the function of the brain when looped with its perceptual organs is not to decode signals, nor to interpret messages, nor to accept

Essa proposta promove uma inovação no olhar sobre a percepção dos seres. Ou seja, considerar a percepção como sentido e não como sensação possibilita encarar o sujeito que percebe como um ser ativo que busca informações em seu ambiente para sobreviver. Essa percepção é real e direta, pois ela não demanda a conversão de dados dos sentidos em percepções. A proposta a que o autor se opõe (a de que os sentidos são considerados como canais de sensações) sugeria que o aprendizado de novos objetos através da percepção ocorria pela anexação de novas respostas a um conjunto fixo de *inputs* perceptuais. No entanto, o autor de *Os sentidos* sugere que “a ênfase é alçada para a descoberta de novos estímulos invariantes, novas propriedades do mundo, para as quais o repertório de respostas da criança pode ser aplicado” (Gibson, 1966: p.5)¹¹¹, i.e., a medida que se amplia a quantidade de invariantes percebidos, amplia-se a capacidade perceptual dos seres.

3.2.2 Como obtemos estímulos?

Diante do fato de que os seres são ativos e equipados com órgãos sensoriais que possibilitam a intercepção de informações invariantes em seu meio de sobrevivência, cabe definir como se obtêm os estímulos do seu nicho. Para responder a essa questão, antes de tudo é preciso definir como o ambiente se apresenta como sendo uma fonte de estimulação para os sentidos.

Gibson propõe que o que há para ser percebido é o ambiente, o meio no qual os seres vivem. Os sentidos, por sua vez, só funcionam quando são estimulados. Assim, para que os seres percebam através dos sentidos o seu meio é logicamente esperado que o ambiente terrestre (composto por elementos sólidos, gasosos e líquidos) seja a fonte de estimulação dos sentidos. Cada um desses elementos geram características em seu meio que colaboram para a promoção da sobrevivência das espécies que aprendem a percebê-las.

images. These old analogies no longer apply. The function of the brain is not even to *organize* the sensory input or to *process* the data, in modern terminology. The perceptual systems, including the nerve centers at various levels up to the brain, are ways of seeking and extracting information about the environment from the flowing array of ambient energy” (Gibson, 1966: p.5).

¹¹¹ “On the assumption that the senses are perceptual systems, however, the emphasis is shifted to the discovery of new stimulus invariants, new properties of the world, to which the child’s repertory of responses can be applied” (Gibson, 1966: p.5).

Por exemplo, a rigidez do ambiente terrestre proporciona um suporte para a sobrevivência, pois “a rigidez dá permanência geométrica aos lugares constância de formas e tamanhos às coisas. Portanto isso 'suporta' não apenas a postura ereta mas também a locomoção, orientação e manipulação” (Gibson, 1966: p.9)¹¹². Soma-se a isso as consequências da gravidade sobre os corpos dos animais que são pressionados à superfície rígida da terra. A gravidade faz com que eles toquem a superfície e ela os toque também. Em outras palavras, a gravidade puxa o animal para baixo e faz com que se produzam duas consequências: 1) há a tensão dentro do corpo do animal; e 2) ocorre a deformação nas partes inferiores em contato com o chão. Diante disso, essas duas consequências gravitacionais fazem com que os animais precisem agir contra essa força para que se mantenham em movimento e eretos. Esse estágio de alerta (movimentação e postura ereta) mostra que o animal está acordado, receptivo e preparado para sobreviver, pois um corpo que não “aguenta” a força da gravidade (um corpo caído ou caindo) se movimenta de maneira distinta. Em suma, perceber essas características da ação da gravidade sobre os corpos promove a sobrevivência.

Da mesma forma, há consequências da radiação eletromagnética, pois é a radiação que mantém o planeta vivo. A radiação, é claro, não é constante e faz com que haja ciclos de atividade tanto químicas quanto de consciência dos animais. São essas diferenças diárias de incidência dos raios solares fazem com que os animais se adaptem a elas para sobreviver.

Ainda mais, é de suma importância para a sobrevivência levar em conta as consequências especiais da luz sobre a vida terrestre. Basicamente, a sua função ecológica é a de iluminar o ambiente e os seres. Contudo, é interessante observar que, para a proposta de Gibson, há uma certa ótica ecológica (*ecological optics*) no tratamento da luz. O tipo de iluminação que incide sobre os seres e o ambiente é chamado de *luz ambiente*, i.e., isso significa que, de maneira geral, os raios de luz convergem sobre um dado corpo. A sua função é a de iluminar os seres e tornar visível a sua posição com relação ao horizonte, fazendo-o ereto ou deitado para um observador, por exemplo. Em oposição a essa luz que se refere aos organismos, há

¹¹² “Rigidity gives geometrical permanence to places and constancy of shape and size to things. It therefore 'supports' not only upright stance but also locomotion, orientation, and manipulation” (Gibson, 1966: p.9).

a *luz radiante*, que é a luz que provém de uma fonte de energia. Esse tipo de luz não é de interesse imediato para a abordagem ecológica promovida aqui uma vez que a sua preocupação é considerar os elementos do meio que promovem a estimulação dos receptores que detectam informações sobre os objetos ao redor do observador e isso só é possível se os objetos forem iluminados, ou seja, há o interesse pela *luz ambiente*.

Por fim, um elemento fundamental para a estimulação dos sentidos dos seres é ar. O meio terrestre é “preenchido” por ar e isso possibilita não só o desenvolvimento da aquisição do ar para a respiração, como também um novo modo de locomoção. Em suas palavras:

“O ar é um excelente meio para a locomoção terrestre. Sendo sem resistência em comparação com a água, isso não requer a anatomia simplificada necessária para um movimento rápido de um peixe, e assim faz-se possível vários experimentos evolucionários com a anatomia dos membros. A adaptação de maior sucesso dos órgãos locomotores foi a que os primatas, que evoluíram mãos ao invés de patas, promovendo a base para um comportamento ainda mais complexo de manipular objetos” (Gibson, 1966: p.14)¹¹³.

Além de possibilitar a evolução da locomoção dos seres que habitam o ambiente terrestre, o ar também é importante devido a possibilidade de transmitir informações nele. A atmosfera é em realidade um meio que possibilita a transmissão de informações de maneira ágil. Para a teoria proposta em *Os sentidos*, a existência de um meio que possibilita tal transmissão é fundamental para amparar proposta de que o ambiente estimula os órgãos dos sentidos. Essa estimulação é possível devido a algumas propriedades existentes no ar, ou seja,

“ele também permite o *fluxo de informação*. Ele permite o fluxo de luz, ele transmite vibração, e ele faz a mediação da difusão de substâncias voláteis. Apenas pela iluminação os animais 'vêem' coisas, apenas pela vibração ele 'ouve' coisas, e apenas pela difusão eles 'cheiram' as coisas. Eu argumentarei que nessas três maneiras várias propriedades das coisas são difundidas, por assim dizer, quando ela é iluminada, ou vibrada, ou

¹¹³ “Air is an excellent medium for terrestrial locomotion. Being unresistant compared to water, it does not require the streamlined anatomy needed by fish for rapid movement, and therefore makes possible various evolutionary experiments with the anatomy of limbs. A most successful adaptation of locomotor organs was that the primates, who evolved hands instead of forepaws, providing a basis for the still more complex behavior of manipulating objects” (Gibson, 1966: p.14).

quimicamente volatizada” (Gibson, 1966: p.14)¹¹⁴.

Como se pode notar, a proposta de Gibson é a de que esses elementos (sólidos, líquidos e gasosos) que compõe o ambiente terrestre promovem a transmissão das informações que caracterizam os eventos ocorridos. Por exemplo, o ar carrega informações dos sons dos animais, pois, de modo geral, os animais são seres que possuem ao seu redor uma certa zona de vibração que caracteriza (sonoramente) a sua ação durante um certo período do dia. Na mesma forma, o ar transmite as informações sobre as substâncias voláteis existentes nos seres, i.e., há a percepção dos odores que podem marcar territórios ou períodos de acasalamento dos animais.

Essa proposta sugere uma certa unificação sobre como percebemos os eventos ao nosso redor, pois possibilita o tratamento simultâneo dos sentidos que possuímos. Para a teoria, o contato com múltiplos sentidos dá mais garantia sobre as propriedades dos objetos e, por sua vez, quanto mais atenção e certeza se tiver sobre os eventos do ambiente, mais a sobrevivência é garantida.

A novidade da proposta de Gibson reside nesse olhar ecológico sobre os seres e os seus respectivos ambientes. Para ele, “o ambiente consiste de *oportunidades* de percepção, de informação *disponíveis*, de estímulos *potenciais*” (Gibson, 1966: p.23)¹¹⁵. E é esse ambiente rico em informações que possibilita o desenvolvimento de ambientes sociais que promovem o desenvolvimento da fala ou de sistemas simbólicos (ou de codificação) que permitem que o homem fale do mundo e torne as palavra os seus equivalentes, por exemplo. Em suma, para Gibson, não há diferenças em se perceber os eventos do ambiente ou perceber a transmissão de informações codificadas sobre o mundo. Em ambos os casos é preciso que ocorra uma transmissão de informações através da vibração do ar, projeção e reflexão da luz ou contato mecânico. Ou seja, ele argumenta que há passos a serem percorridos para a transmissão de informação. Parte-se da expressão vocal, passa-se pela expressão gestual e facial, depois à realização de

¹¹⁴ “It also permits the *flow of information*. It permits the flux of light, it transmits vibration, and it mediates the diffusion of volatile substances. Only by illumination do animals 'see' things, only by vibration do they 'hear' things, and only by diffusion do they 'smell' things. I shall argue the in these three ways various properties of a thing are broadcast, as it were, when it is illuminated, or vibrating, or chemically volatile” (Gibson, 1966: p.14).

¹¹⁵ “The environment consists if *opportunities* for perception, of *available* information, of *potential* stimuli” (Gibson, 1966: p.23).

imagens e figuras e , por último, a escrita. Em todas essas etapas, a transmissão de informação só é possível por serem eventos no ambiente que estimulam os órgãos receptores dos seres.

Diante disso, chega o momento de respondermos: mas como os sentidos obtêm as informações a partir dos estímulos do ambiente? Para responder a essa pergunta, Gibson propõe que a percepção é baseada na atividade dos receptores, dos órgãos e dos sistemas perceptuais.

Os receptores são as unidades mais baixas (microscópicas) utilizadas para a percepção. Basicamente, trata-se de células que respondem às energias química, mecânica ou luminosa aplicadas a elas. O que deve ser levado em conta nesta proposta é os receptores (as células) são estruturas funcionais. Concebiam-se as células como estruturas anatômicas que compunham os órgãos do sentido. No entanto, há evidências que mostram que há células que fazem parte de determinados órgãos que só respondem a certos tipos de estímulos, indicando que somente são ativadas para cumprir determinadas funções. Diante disso, Gibson desvincula a noção de que as células são receptores de energia que transmitem levam estímulos ao cérebro pelos nervos, Para o autor,

“Essas unidades parecem modificar seus inputs como uma função de mudança de energia (ordem sequencial) ou uma relação de energia (ordem adjacente), não como uma função de aplicação de energia. O que elas registram não é a energia de estimulação, mas a sequência ou o arranjo da estimulação, ou seja, a informação” (Gibson, 1966: p.41)¹¹⁶.

Por sua vez, para definir o que são os órgãos, Gibson comenta que o termo receptor é ambíguo, pois ele pode ser entendido como um mecanismo passivo que dá início aos impulsos nervosos quando “tocado” por um estímulo ou como uma estrutura ativa que se ajusta por conta própria a partir das possibilidades de estimulação. Na sua teoria, os mecanismos passivos e estáticos são os receptores, as células. Já os órgãos seriam essas estruturas ativas compostas por receptores, músculos e que devem ser supridas por muitas fibras nervosas que conduzem os estímulos ao sistema nervoso central.

¹¹⁶ “These units seem to modify their input as a function of a change of energy (sequential order) or a relation of energy (adjacent order), not as a function of the application of energy. What they register is not the energy of stimulation, but the sequence or arrangement of stimulation, that is, information” (Gibson, 1966: p.41).

Contrariamente ao senso comum, a proposta de *Os sentidos* diz que não há um relação biunívoca entre um órgão, um tipo de receptor e, conseqüentemente, um nervo. Segundo Gibson (1966: p.42), os órgãos estão organizados hierarquicamente, estando os menores subordinados aos maiores. Essa organização promove uma integração entre os órgãos e faz com que os sentidos não correspondam especificamente a certos nervos.

Ainda pensando sobre os órgãos, vale salientar que eles podem obter estimulação de duas maneiras. O autor caracteriza alguns órgãos como sendo tipicamente motores, pois executam ações, e outros, como sensoriais, porque recebem estímulos do ambiente de modo a explorá-lo e a investigá-lo. No entanto, essa caracterização não é exclusiva. A mão, por exemplo, é um órgão que tanto performa ações quanto explora o ambiente em busca de informações. Isso mostra que essas duas maneiras de obter estimulação são integradas pela atividade dos observadores que são ativos, i.e., eles buscam informações em seu ambiente.

Por fim, os sistemas perceptuais são constituídos dos receptores e órgãos. Por exemplo, o sistema visual possui vários níveis. Ao se levar em conta um único olho, temos um sistema de baixo nível, pois se trata apenas de um órgão (o olho) e as suas respectivas células. No entanto, se considerarmos esse olho aliado a uma atividade muscular, teremos um sistema de alto nível, porque ele estará estabilizado na cabeça e, com o auxílio do ouvido, conseguirá escanear o ambiente em busca de informações. Por sua vez, se considerarmos os dois olhos juntos compondo um sistema visual de alto nível, teremos os dois olhos alocados na cabeça e atrelados a um sistema corporal, que coopera com a locomoção e com o equilíbrio postural do corpo, fazendo com que o homem consiga olhar tudo que está ao seu redor. Em suma, um sistema perceptual leva em conta não só os estímulos impostos às células e obtidos pelos órgãos, ele trata das informações dos eventos perceptuais de maneira distinta, pois se trata de um sistema de alta ordem.

A proposta de Gibson é tratar não só da atividade das células ou dos órgãos, ele pretende considerar a percepção a partir dos sistemas perceptuais, i.e., trata-se da integração de vários níveis (e de sistemas também) que detectam as informações do ambiente que promovem a percepção dos eventos. Na próxima seção (3.2.3 do Capítulo 2), trataremos mais especificamente do funcionamento dos sistemas

perceptuais.

3.2.3 Os sistemas perceptuais

Gibson comenta que há a necessidade de se propor uma nova terminologia para o tratamento da percepção, uma vez que os sentidos são tidos

“como ativos ao invés de passivos, como sistemas ao invés de canais e como inter-relacionados ao invés de mutualmente exclusivos. Se eles funcionam para coletar informação, não simplesmente para emergir sensações, essa função deveria ser denotada por um termo diferente. Aqui eles serão chamados de *sistemas perceptuais*” (Gibson, 1966: p.47)¹¹⁷.

Em outras palavras, os sistemas perceptuais (tal como apresentado anteriormente na seção 3.2.3 do Capítulo 2)) são compostos pelos receptores e órgãos dos organismos e são hierarquicamente organizados. Isso não significa que um sistema é superior ao outro, pois a hierarquização é dado dentro de cada sistema e não entre sistemas. No entanto, resta-nos perguntar quais seriam esses sistemas e como eles funcionariam?

Antes de responder a essa questão, Gibson comenta que as doutrinas anteriores acerca da percepção tentavam elaborar um catálogo de sensações. No entanto, como se tentava atrelar um sentido a uma sensação e uma sensação a um nervo específico que transmitiria sensações ao cérebro, essa empreitada falhou. Como se pode ver, esse catálogo não funcionava na sua própria origem, pois quais seriam as sensações transmitidas pela pele?; qual o nervo específico que transmite as informações?; e para a gustação?; Qual seria o nervo gustatório? Ainda, a construção desse catálogo de sensações falhava em considerar estimulações ativas como passivas. Por exemplo, a obtenção de estímulos com o auxílio de um órgão como a mão (ou através do tato, como queria a abordagem por sensações) é antes de tudo exploratória. A mão busca informações no ambiente. Em outras palavras, é melhor tratar os sentidos não como sensações e sim como modos de atenção externa, i.e., os órgão que compõem os sistemas perceptuais são ativos e buscam

¹¹⁷ “We shall have to conceive the external senses in a new way, as active rather than passive, as systems rather than channels, and as interrelated rather than mutually exclusive. If they function to pick up information, not simply to arouse sensations, this function should be denoted by a different term. They will here be called *perceptual systems*” (Gibson, 1966: p.47).

informações no ambiente.

Diante disso, Gibson sugere a existência de cinco modos de atenção externa. Essa sua classificação, ele está usando como critério os modos de atividade que coletamos informações sobre o ambiente e não os modos de qualidade consciente, ou seja, há aqui uma classificação que desvincula a detecção de informação com as “possíveis” sensações. Sua proposta caminha na direção de dizer que não relação de causalidade entre coleta de informação e sensação, uma vez que pode haver percepção sem emergir sensações.

Basicamente, são cinco os sistemas perceptuais: orientação; audição; tato; gustativo-olfativo; e visão. Algum desses sistemas podem coletar as mesmas informações que outros sistemas. Essa redundância na coleta promove a cooperação entre os sistemas nas variadas combinações. Para Gibson, esses sistemas devem funcionar para o homem, outros mamíferos e a maioria dos vertebrados. Esse ponto é claramente uma preocupação ecológica da proposta, que tem por objetivo sugerir que a percepção não é algo especial dos seres humanos. Esse ponto será aproveitado pela TRD em oposição à abordagem especializadora da TM. No Tabela 1, reproduzimos a classificação de Gibson para os sistemas perceptuais bem como o detalhamento de cada um.

Por exemplo, o sistema de orientação básica trabalha com a coleta de informações sobre as forças de aceleração através do ouvido interno. Tais dados especificam a direção da gravidade que determina o início e o término dos movimentos do corpo devido a sua constante atuação sobre ele. Ainda mais, o *input* sobre a gravidade é duplamente registrado pela cooperação do sistema tátil, pois “sente na pele” a sua atuação.

Já o sistema auditivo especifica a natureza dos eventos vibratórios do mundo e, quando ou dois ouvidos são usados para explorar o ambiente, é possível identificar a sua direção. Por sua vez, o sistema tátil é composto de um complexo emaranhado de subsistemas. Em realidade, não há órgãos do sentido no sentido convencional do termo, mas os receptores epiteliais estão em quase todos os lugares e os receptores das articulações cooperam para a obtenção de informações sobre o mundo. Na visão de Gibson, o sistema tátil é poderoso devido a sua

atividade de exploração, mas como a maioria dos seres depende da visão pouco se sabe sobre a sua capacidade de obter informações. O autor somente comenta que o tato e a visão frequentemente fornecem informação redundante, garantindo um *input* duplo de informação.

Nome	Modo de atenção	Unidades representativas	Anatomia do órgão	Atividade do órgão	Estímulo disponível	Informação externa obtida
Sistema básico de orientação	Orientação geral	Receptores mecânicos	Órgãos vestibulares	Equilíbrio do corpo	Força da gravidade e aceleração	Direção da gravidade, ser empurrado
Sistema auditivo	Ouvir	Receptores mecânicos	Órgãos cocleares como o ouvido médio e o aurículo	Orientação para os sons	Vibração do ar	Natureza e localização dos eventos vibratórios
Sistema tátil	Tocar	Receptores mecânicos e possivelmente térmicos	Pele e juntas	Exploração de várias maneiras	Deformação dos tecidos, Configuração das juntas, fortalecimento das fibras musculares	Contato com a terra, encontros mecânicos, formas dos objetos, esta dos dos materiais (sólidos ou viscosos)
Sistema gustativo-olafativo	Cheirar	Receptores químicos	Cavidade nasal	Farejar	Composição do meio	Natureza das fontes voláteis
	Sentir gosto	Receptores químicos e mecânicos	Cavidade oral	Saborear	Composição dos objetos ingeridos	Valores nutritivos e bioquímicos
Sistema visual	Olhar	Fotorreceptores	Mecanismo ocular	Acomodação, ajustes pupilares, fixação, exploração convergente	As variáveis de estrutura no ambiente iluminado	Tudo que pode ser especificado pelas variáveis da estrutura óptica (informação sobre objetos,

						animais, movimentos, eventos e locais.
--	--	--	--	--	--	---

Tabela 1: Os sistemas perceptuais e seu detalhamento (Gibson, 1966: p.50)

O sistema gustativo-olfativo ilustra satisfatoriamente como há a integração de outros sistemas na execução de certas tarefas. Embora se pense que esse sistema faça uso apenas de dois órgãos distintos (o nariz e a boca), o ato de sentir um alimento na boca envolve o sistema tátil. Esse envolvimento tátil é fundamental para que se realizem “testes” químicos e físicos antes de engolir a comida. Tais testes determinam o cheiro do alimento (através das substâncias voláteis que o compõem), sua solubilidade e consistência física. Por fim, o sistema visual se coloca como um de muitas capacidades, pois ele consegue detectar cores e movimentos dos eventos até mesmo à longa distância. Entretanto, isso só é possível se tais eventos estiverem iluminados.

Em suma, cada um dos sistemas pode obter informações sobre o ambiente por si só ou com o auxílio de outros sistemas. Isso não sugere, tal como já foi mencionado anteriormente, que não há uma “superioridade” de um sistema sobre o outro. Há, certamente, especializações dos sistemas em detectar certos tipos de informações, mas não há a necessidade que um outro sistema confirme a informação detectada.

Por exemplo, perceber o fogo envolve um conjunto de sensações sonoras, olfativas, táteis (calor) e visuais (coloração). Todas elas têm de ser associadas a uma experiência passada que explicaria todos esses elementos a partir da memória. Assim, não importaria qual estímulo estaria disponível num dado momento que se perceberia o fogo, pois, estando todos os sistemas ativos, a informação sobre o fogo seria redundante.

Quando se fala na relação da percepção com uma experiência passada, é preciso esclarecer que

“nesta teoria o problema da percepção não é como as sensações são associadas; é como o som, o odor, o calor, ou a luz que especifica o fogo é discriminado de

todos os outros sons, odores, calores, e luzes que não especificam o fogo” (Gibson, 1966: p.55)¹¹⁸.

Como se pode ver, a teoria exposta em *Os sentidos* não está preocupada em determinar como os eventos do ambiente são estocados na memória. Trata-se de uma teoria que busca traçar quais são os invariantes que são detectados pelos sistemas perceptuais. A esta altura, a pergunta a ser respondida é: o que é inato e o que é adquirido na percepção? Para responder a esta questão, Gibson determina as bases do que ele chamou de Teoria da Coleta de Informação. Trataremos dela na próxima seção (3.2.4 do Capítulo 2).

3.2.4 A teoria da coleta de informação

Na trajetória da apresentação da sua teoria de percepção, Gibson tenta deixar claro que sua proposta se opõe às teorias baseadas na sensação ([theories of sensation-based perception]). Nessas teorias, a percepção depende totalmente das sensações que são específicas dos receptores. A sua teoria é de outra natureza. Ele a considera uma abordagem baseada na informação. Em suas palavras:

“A presente teoria afirma que a possibilidade da experiência perceptual sem qualidades sensoriais subjacentes que são específicas dos receptores, e eu a tenho chamado de uma teoria de percepção baseada na informação. Isso é uma nova orientação” (Gibson, 1966: p.266)¹¹⁹.

A novidade que orienta a teoria proposta em *Os sentidos* (uma teoria baseada na informação) tenta dar outro direcionamento sobre a antiga controvérsia entre empiristas e inatistas. Enquanto estes postulavam que os seres possuem uma mente que apresenta capacidades inatas mesmo antes de perceber os objetos do mundo, aqueles afirmam que os seres começam a vida sem nenhum conhecimento na mente. Para os adeptos do arcabouço empirista, apenas há a capacidade de se perceber as sensações sem sentido algum e somente se aprende a perceber o mundo através da memória e da associação após o acúmulo de experiências

¹¹⁸ “In this theory the problem of perception is not how sensations get associated; *it is how the sound, the odor, the warmth, or the light that specifies fire gets discriminated from all the other sounds, odors, warmths, and lights that do not specify fire*” (Gibson, 1966: p.55).

¹¹⁹ “The present theory asserts the possibility of perceptual experience without underlying sensory qualities that are specific to receptors, and I have called this a theory of information-based perception. It is a new departure” (Gibson, 1966: p.266).

passadas.

Claramente, cada uma das propostas salienta aspetos distintos envolvidos na percepção do mundo. O inatismo traz à luz a hereditariedade do seu desenvolvimento e o empirismo enfoca a influência do aprendizado. No entanto, nenhum dos pontos de interesse dessas abordagens são relevantes para a proposta gibsoniana, uma vez que por mais que já se tenha algum conhecimento antes mesmo do nascimento, é preciso que os sistemas perceptuais sejam estimulados pelo ambiente para que se perceba o que se está no meio. Da mesma forma, em resposta aos empiristas, os sistemas perceptuais não precisam lidar com a formação de associações e a suas relações com a memória. Isso não é preciso porque os sistemas perceptuais buscam ativamente por informações estruturadas no ambiente. Não se trata de se buscar pelos elementos pré-determinados inatamente que compõem as coisas do mundo ou de se absorver tudo dela. Trata-se de se perceber as estruturas do mundo, uma vez que o mundo é estruturalmente organizado. Esse ponto fica mais evidente quando o autor apresenta a sua definição do que é mente. Em suas palavras:

“o cérebro pode ser tratado como o mais alto dos vários centros do sistema nervoso que comanda os sistemas perceptuais. Ao invés de postular que o cérebro constrói informações a partir do input de um nervo sensorial, nós podemos supor que os centros do sistema nervoso, incluindo o cérebro, ressoam a informação” (Gibson, 1966: p.267).¹²⁰

Como se pode notar, a sua proposta teórica coloca a mente como um sistema que comanda os sistemas perceptuais, mas que está atento às informações do ambiente. Ainda mais, o cérebro ressoa as informações encontradas no ambiente e não uma representação ou determinada *a priori* ou puramente aprendida. A teoria sugerida por Gibson, em realidade, não consegue responder quais mecanismos de detecção de informação aparecem no nascimento e quais dependem do aprendizado, porque não haveria mecanismos puramente genéticos ou aprendidos. Existe sim mecanismos intermediários que estão relacionados à maturação e ao crescimento dos seres:

¹²⁰ “The brain can be treated as the highest of several centers of the nervous system governing the perceptual systems. Instead of postulating that the brain constructs information from the input of a sensory nerve, we can suppose that the centers of the nervous system, including the brain, resonate information” (Gibson, 1966: p.267).

“A anatomia e a fisiologia básica dos órgãos da percepção dependem principalmente dos fatores genéticos tal como determinados pela evolução. A maturação dos sistemas perceptuais dependem dos determinantes genéticos e ambientais combinados. A educação dos sistemas perceptuais depende principalmente da história da exposição do indivíduo ao ambiente” (Gibson, 1966. p. 268)¹²¹.

Pensando nisso, surgem três questões: 1) quanto da percepção depende dos órgãos?; 2) quanto da percepção depende do crescimento?; e 3) quanto da percepção depende da experiência?

Dentro da perspectiva evolucionista que Gibson trabalha, a primeira questão é respondida considerando que os órgão envolvidos na percepção (tal como a consideramos neste período da história dos seres) foram moldados através dos anos para responder às necessidades ecológicas. Os órgãos seriam estruturas especializadas para algumas tarefas (como ouvir, cheirar, ver, tocar e saborear), mas não são para outras. Soma-se a isso o fato de que as informações registradas pelos sentidos são, em grande parte, redundantes. Tais aspectos mostram que há modos de atenção (como orientação e exploração) que são parcialmente restringidos pela anatomia e, da mesma maneira, parcialmente livres para se adaptarem à outras atividades¹²².

A resposta para a segunda e a terceira perguntas estão relacionadas, porque para Gibson a habilidade de perceber melhora a medida que os seres crescem. Há em realidade um processo de maturação dos órgãos e da atenção sobre o seu ambiente. Sua perspectiva teórica não vê a criança como um adulto em miniatura (*um adulto sem experiência* ou *uma alma sensível sem memória*), pois “a habilidade de selecionar e abstrair informações sobre o mundo cresce a medida que ele o faz” (Gibson, 1966: 269)¹²³, i.e., a criança perceberá aquilo que ela será capaz de perceber a medida que sua capacidade de abstração e atenção forem se

¹²¹ “ The anatomy and basic physiology of the organs of perception depend mainly on genetic factors as determined by evolution. The maturation of the perceptual systems depends on genetic and environmental determiners in concert. The education of the perceptual systems depends mainly on the individual's history of exposure to the environment” (Gibson, 1966: p.268).

¹²² Um exemplo disso seria a especialização que o aparelho fonador dos seres humanos assume, pois há nele a combinação de funções como alimentação, respiração e fonação. Da mesma forma, “livremente” ele se adaptou para se estabelecerem diferenças de pressão intraorais para se tocar trompetes e trombones (cf. Ohala, 1996).

¹²³ “The ability to select and abstract information about the world grows as he does” (Gibson, 1966: p.269).

desenvolvendo ou maturando. Isso é possível pois os seres nunca param de experienciar coisas e, conseqüentemente, aprendê-las. Podemos verificar isso ao pensarmos que, embora a detecção de informações pode ser afetada quando os receptores “envelhecem” (e.g., quando as lentes dos olhos e os ossos perdem a sua “flexibilidade juvenil”), músicos e artistas conseguem descobrir novas variações de luz e de som com o passar do tempo.

Como se pode notar, as três respostas sugerem que a percepção depende desses três fatores: órgãos moldados pela evolução para cumprir tais funções e se adaptarem a novas; maturação das habilidades perceptuais vão sendo adquiridas a medida que os seres crescem; e, da mesma maneira que crescemos e experienciamos novos eventos, aprendemos mais.

Assim, o autor sugere que é preciso um mecanismo para aprender a perceber, pois aprendemos a perceber enquanto experienciamos eventos. Esse mecanismo aprende novas distinções a partir da apresentação de novos padrões estruturados do ambiente. A regra de aprendizagem dos sentidos é *aprender*. Gibson (1966: p.270) comenta que há experimentos psicofísicos que mostram que há diminuição na taxa de erros de discriminação quando os informantes não recebem retorno se estão errando ou acertando. Para ele, isso é um indício de que os sistemas perceptuais somente seguem a regra de *aprender*, indicando também que os seres “educam a sua atenção” para a informação disponível nos estímulos. Em resumo, a atenção cresce com a prática.

Gibson & Gibson (1955 *apud* Gibson, 1966: p.270) propõem a chamada *teoria da diferenciação* para dar conta do fato de os seres detectarem os traços distintivos dos objetos e eventos de seu ambiente e abstraírem suas propriedades gerais. Essa proposta sugere que “um percebedor é um sistema autoajustável”, pois a ação do seu sistema nervoso, em realidade, está ocupada com a ressonância das informações dos estímulos. Em outras palavras, a ação do percebedor não é a de armazenar no cérebro imagens ou conexões das células nervosas, e sim a de se reforçar a coleta de informações do ambiente até o sistema obter “clareza”, i.e., não se trata da clareza disponibilizada pela luz, trata-se de o sistema saber que ele ressoou o estímulo buscado. Um exemplo disso é o autoajuste das pupilas dos olhos a partir da variação de luminosidade disponível ao percebedor.

Como podemos notar, a proposta da teoria da coleta de informações lança um novo olhar sobre como percebemos o mundo e sobre como aprendemos a percebê-lo. Ela deixa de lado as visões associativas acerca dos eventos disponíveis no ambiente e tenta desvincular a memória dos processos perceptuais.

Grosso modo, quando se fala de associações em *Os sentidos*, as consideramos como *associações ecológicas*. Isso significa que só há associação quando há “invariância na combinação de estímulos”. Como exemplo, podemos citar a percepção do que é um pêssego. Ele só será percebido como um pêssego se esse objeto sempre fornecer uma certa coloração, odor, textura e sabor característicos. Esses traços devem ser todos típicos de um mesmo objeto e devem formar uma única combinação. Nas palavras de Gibson:

“o ato de perceber fogo ou pêssego, então, poderia muito bem ser considerado a coleta das variáveis associadas da informação como os dados sensoriais de associação. Duas coisas são necessárias: as dimensões de qualidade devem ter sido diferenciadas, e as combinações invariantes de qualidade devem ser detectadas. A formação de associações não é necessária” (Gibson, 1966: p.272)¹²⁴.

Como se pode ver, a sua perspectiva do que é uma associação é distinta da proposta por Skinner (1938 *apud* Gibson, 1966: p.272), porque não há a proposição de condicionamento de ações e tampouco de controle sobre situações experimentais. A perspectiva de Gibson trata do aprendizado da percepção em ambientes não controlados ou naturais. Em suma, ao invés de propor um aprendizado *por associação* (tal como queria a perspectiva behaviorista), ele sugere um *de associações*, i.e., o que interessa são as associações (de estímulos invariáveis a um dado objeto) percebidas e aprendidas.

Por sua vez, ao pensar na relação da percepção com a memória, Gibson combate a proposta de Lashley (1950 *apud* Gibson, 1966: p.275) de que haveria uma base fisiológica para o lembrar de coisas. A tentativa de se determinar um local onde se armazenam as lembranças falhou e a hipótese gibsoniana de que o cérebro se autoajusta e ressoa as informações coletadas passa a dar um tratamento distinto

¹²⁴ “The act of perceiving a fire or a peach, then, might just as well be considered the pickup of the associated variables of information as the associating of sensory data. Two things are necessary: the dimensions of quality must have been differentiated, and the invariant combinations of quality must be detected. The formation of associations is not necessary” (Gibson, 1966: p.272).

para as concepções tradicionais acerca da memória.

O autor comenta que tradicionalmente se vincula a memória com o passado, a percepção ao presente e a introspecção analítica ao futuro. Na sua abordagem, a percepção não está no presente e sim a sensação. Em realidade, é difícil determinar onde a memória acaba e a percepção começa, pois a informação disponível no ambiente para ser coletada não existe exclusivamente no passado, presente ou futuro. A confusão que se faz é a de que a experiência do “agora” é resultado da atenção do observador ao seu próprio corpo e às impressões realizadas nele, i.e., estamos falando de sensação e não de percepção.

Em outras palavras, a proposta de Gibson não é de que não lembramos dos eventos percebidos anteriormente. O que ele quer dizer é que a memória não é a base do aprendizado. Os seres aprendem porque eles são sistemas autoajustáveis que buscam informações nos objetos do ambiente:

“Em resumo, o observador humano aprende a detectar o que tem sido chamado de valores ou significados das coisas, perceber seus traços distintivos, colocá-los em categorias e subcategorias, notar suas similaridades e diferenças e ainda estudá-los para seu próprio bem, além de aprender o que fazer com eles. Toda essa discriminação, sendo otimista, tem sido baseada inteiramente na educação da sua atenção sobre as sutilezas da informação invariante dos estímulos” (Gibson, 1966: p.285)¹²⁵.

Para que essa percepção ocorra, Gibson sugere um último ponto importante para o desenvolvimento da percepção. Trata-se da proposta de que é indispensável o desenvolvimento de uma atenção seletiva, i.e., os seres em desenvolvimento aprendem a se focarem apenas nas características distintivas e relevantes para que a percepção dos eventos que ocorrem em seu ambiente seja rápida e eficiente. Essa percepção econômica promoveria, é claro, a sobrevivência de algumas espécies e o fim de outras. Em outras palavras, podemos notar que esse princípio de economia seleciona informações relevantes ao invés de organizá-las:

“[...] apenas as informações requeridas para

¹²⁵ “In short, the human observer learns to detect what have been called the values or meanings of things, perceiving their distinctive features, putting them into categories and subcategories, noticing their similarities and differences and even studying them for their own sakes, apart from learning what to do about them. All this discrimination, wonderful to say, has to be based entirely on the education of his attention to the subtleties of invariant stimulus information” (Gibson, 1966: p.285).

economicamente identificar algo tendem a ser coletadas a partir de uma informação complexa do estímulo. Todas as outras informações disponíveis, que seriam requeridas para especificar sua identidade única e completa em todo universo de coisas, não são levadas em conta” (Gibson, 1966: p.286)¹²⁶.

A proposta de Gibson para o tratamento da percepção se apresenta como uma alternativa promissora para se explicar como se dá a percepção da fala. Ao invés das explicações puramente de base auditiva ou inatistas (como a da TM), a disponível em *Os sentidos* aponta para uma nova organização da mente, dos sentidos e dos dados disponíveis no ambiente. Tais questões lançam luz sobre os primeiros achados acerca da influência de aspectos coarticulatórios sobre a percepção (apresentados da seção 3.1 do Capítulo 2), pois inicia uma reflexão sobre como o sistema de percepção da fala se ajusta ou detecta informações disponíveis sobre as realizações fônicas dos falantes. Na próxima seção (3.3 do Capítulo 2), apresentaremos uma primeira formulação da TRD. Nela, há explicitamente um vínculo epistemológico com a abordagem de base ecológica proposta por Gibson (1966). Vamos a ela!

3.3 Uma primeira formulação da Teoria do Realismo Direto

Nesta seção apresentaremos a primeira formulação da TRD (Fowler, 1986). Embora a autora apresente evidências importantes para um tratamento direto e realista sobre a percepção da fala, vale salientar que ainda se tratava de um programa de pesquisa. Por este motivo, não há, ainda, considerações explícitas acerca dos aspectos multimodais da percepção. Passemos, então, à proposta, às evidências experimentais e ao programa de pesquisa a ser desenvolvido.

3.3.1 Proposta

A proposta de Fowler (1986) para a percepção da fala se apresenta como

¹²⁶ “The rule is, I suggest, that only the information required to identify a thing economically tends to be picked up from a complex of stimulus information. All the other available information that would be required to specify its unique and complete identity in the whole universe of things is not attended to” (Gibson, 1966: p.286).

uma novidade nos estudos linguísticos. O seu tratamento decorre justamente do fato de não haver teorias de percepção que tomem os fenômenos como *eventos ecológicos*. Em suas palavras:

“Um evento ecológico é uma ocorrência em um ambiente definido com respeito aos seus potenciais participantes. Como a maioria dos eventos ecológicos (doravante, eventos), um evento do qual a comunicação linguística faz parte é altamente estruturado e complexo” (Fowler, 1986: p.3)¹²⁷.

Diante disso, os estudos da comunicação linguística tomam muitas formas devido à falta de uma partição natural sobre os fenômenos. É inegável também salientar que as abordagens correntes naquele momento acabavam deixando de lado outros aspectos necessários para a comunicação, como os gestos utilizados pelos falantes, os aspectos presentes no ambiente que possibilitam ao falante apenas apontar ao invés de falar e os próprios interlocutores que podem afetar o estilo de fala do enunciador.

Fowler, então, decide apresentar uma partição sobre os fenômenos da linguagem para que sua abordagem a partir de eventos seja satisfatoriamente acomodada em uma teoria da comunicação. Seu ponto de partida é a proposta de Hockett (1960 *apud* Fowler, 1986: p. 3), na qual o autor sugere que as línguas exibem uma “dualidade de padronização¹²⁸”, ou seja, as línguas possuem palavras organizadas gramaticalmente em sentenças e segmentos fonéticos organizados fonotaticamente em palavras.

Com essa proposta de organização gramatical, é possível conceder dois tipos de poderes às sentenças linguísticas. O primeiro se refere ao conteúdo comunicativo de uma sentença ser maior do que o conteúdo das palavras que a compõem. Segundo, os falantes de uma dada língua podem produzir novas sentenças que a audiência nunca havia ouvido antes. Essa primeira padronização é tomada como evento linguístico por Fowler e ela decide realizar poucos comentários sobre tais

¹²⁷ “An ecological event is an occurrence in the environment defined with respect to potential participants in it. Like most ecological events (henceforth, events), one in which linguistic communication takes place is highly structured and complex” (Fowler, 1986: p. 3).

¹²⁸ Embora Hockett tenha formulado o termo dualidade de padronização, uma ideia semelhante a essa se encontra nos trabalhos de Martinet dos anos 50 com o nome *dupla articulação*. Embora no Dicionário de Linguagem e Linguística de Trask (2004) faz-se menção à “dualidade de esquematização”. Manteremos aqui o termo “dualidade de padronização” para tentar remeter às ideias de Hockett utilizadas por Fowler.

fatos.

A outra parte da padronização das línguas é chamada de eventos da fala e é sobre essa face que o trabalho vai ser desenvolvido. Nessa parte da dualidade das línguas é que se considera, por exemplo, a constituição do léxico, já que as unidades de interesse são os segmentos fonéticos que constituem as palavras.

Com o intuito de encaminhar as suas escolhas no desenvolvimento da teoria, a autora afirma que tais unidades que constituem as palavras têm que ser de base articulatória uma vez que, a partir de simulações em computador realizadas por Lindblom, MacNeilage & Studert-Kennedy (1983 *apud* Fowler, 1986: p.4), o nosso léxico seria muito pequeno se apenas considerar os segmentos organizados fonotaticamente. Tais simulações mostram um aumento surpreendente no tamanho do léxico se levarmos em conta a abertura e o fechamento do trato vocal, por exemplo. Em outras palavras, os autores sugerem que os primitivos fonológicos que compõe as unidades lexicais são de base articulatória e assumi-lo explicaria como ocorre o aumento do léxico.

Diante disso, “um evento de fala também pode ser definido como uma sentença linguística que tem estruturas fonéticas percebidas por ouvintes” (Fowler, 1986: p.4) e é a partir da instanciação das trocas entre falantes e ouvintes que se argumenta em favor de um tratamento realista e direto para a percepção. Para a autora, a maneira de explicar o sucesso da comunicação entre usuários reais da língua só é possível com a adoção de uma teoria que considere que as atividades dos sujeitos sejam guiadas perceptualmente, ou seja, que se recuperem os eventos do mundo real.

Em seu trabalho, a autora assume a partição proposta por Hockett para dizer que vai se deter às estruturas fonéticas sem necessariamente se importar com os significados gramaticais das sentenças. Em suma, a autora defende “que os falantes produzem fala foneticamente estruturada, os ouvintes a percebem como tal e eles usam a estrutura fonética percebida para guiar seu comportamento subsequente” (Fowler, 1986: p.4). A sua proposta não é a de que os eventos da fala sejam primários ou privilegiados. Ela apenas os está defendendo como parte integrante de uma comunicação linguística e que, por sua vez, deve ser contemplado em uma

teoria de percepção.

Assim, a sua proposta de percepção requer três elementos para que se obtenha sucesso comunicativo: os eventos do ambiente ou eventos distais; uma mídia informacional; e os percebedores. Os eventos do ambiente não são todos os eventos disponíveis. Na percepção da fala, é preciso considerar os eventos distais localmente, uma vez que o evento de interesse para os percebedores é o trato vocal articulado. As informações sobre os eventos distais serão estruturadas em uma mídia informacional, que no caso da fala é o ar e no da visão, a luz refletida, por exemplo. Essa mídia, então, transmite informações através da estimulação dos órgãos do sentido de um percebedor sensível a essa mídia e, é claro, a essas informações.

Vale salientar ainda que esses percebedores são ativos, i.e., eles procuram a informação relevante para suprir suas necessidades e interesses. A noção de percebedor ativo recebe duas interpretações. A primeira se refere ao fato de que eles se movimentam no ambiente para interceptar essas informações relevantes. Já a segunda está relacionada ao aspecto de que os percebedores sintonizam os seus sistemas perceptuais para responder seletivamente a tais aspectos disponíveis no ambiente.

Pensando, então, localmente na percepção da fala, temos segmentos fonéticos realizados como gestos coordenados a partir das estruturas do trato vocal (eventos distais) que são veiculados em um meio acústico (meio informacional) para estimularem o sistema perceptual de percebedores envolvidos em um processo comunicativo de um dado ambiente¹²⁹.

A partir dessa proposta, a autora não apresenta diretamente argumentos em favor da teoria. Ao invés disso, a sua estratégia é apresentar quatro barreiras que devem ser superadas. Trata-se de uma “agenda” para o seu desenvolvimento e ela versa basicamente sobre o fato de que para fundamentar satisfatoriamente a proposta realista e direta da percepção, antes de tudo devem ser descritas as

¹²⁹ A assunção de que os segmentos fonéticos são gestos coordenados a partir de estruturas do trato vocal será adotada pela Fonologia Articulatória (Browman & Goldstein, 1986 e seguintes). Para endossar a sua proposta gestual da percepção da fala, Fowler & Rosenblum (1991: p.48) comentam os avanços promovidos pela proposta gestual de Browman & Goldstein e pela implementação do gesto articulatório no modelo de Saltzman (1986) e Saltzman & Kelso (1987).

realizações articulatórias dos segmentos fonéticos e identificadas as suas consequências acústicas para, em seguida, estudar o uso das estruturas que guiam as ações dos ouvintes.

Passemos, então, às quatro barreiras a serem superadas pela abordagem realista direta da percepção da fala.

3.3.2 Quatro barreiras para o Realismo Direto

3.3.2.1 *Se os ouvintes recuperam a articulação, por que eles não sabem disso?*

A proposta de que os ouvintes percebem as estruturas articulatórias coordenadas do trato vocal ao invés do som enfrenta maior resistência do que a mesma abordagem para a percepção visual. Essa resistência está relacionada às concepções tradicionais do que seriam os segmentos fonéticos. Por exemplo, Repp (1981 *apud* Fowler, 1986: p.6) argumenta que os segmentos fonéticos são “abstrações” e produtos dos processos cognitivos aplicados à estimulação dos sentidos. Para ele, os segmentos não têm propriedades físicas como a duração, espectro ou amplitude, e, por isso, não são passíveis de medições.

Essa questão fica mais resistente ainda quando pensamos que no momento em que os usuários de uma língua ouvem a fala não estão conscientes que estão ouvindo o fechamento de lábios ou a abertura de mandíbula. No entanto, é preciso ter em vista que também não estão experienciando traços do sinal acústico.

O fato é que a autora não consegue explicar o motivo pelo qual a nossa intuição falha ao se pensar quais são os primitivos da nossa percepção. No entanto, ela argumenta que, se há evidências experimentais de que os usuários de uma língua seguem características articulatórias, parece haver um descompasso entre as descobertas oferecidas pelos experimentos de percepção e o conhecimento que se tem sobre as articulações da fala.

Diante disso, Fowler apresenta alguns dados que evidenciam que a percepção da fala é guiada por um conjunto de movimentos coordenados das

estruturas do trato vocal.

Um primeiro exemplo se refere ao fato de que, em contextos artificiais, um breve silêncio no sinal pode promover a percepção de uma oclusiva. Em contrapartida, em ambientes reais de fala, uma breve oclusão entre um [s] e a sílaba [lit] não produzem uma consoante como [p], mesmo se as transições de [p] estiverem disponíveis no sinal. Tais estudos mostram que essas pistas acústicas (oclusão e transições) podem cooperar ou conflitar. As pistas cooperam quando os dois segmentos da sílaba apresentam evidências para a produção da oclusiva, i.e., “quantidade” ideal de silêncio e transições específicas de um [p]. Elas entram em conflito quando alguma das pistas não está presente ou não é suficiente para promover a sua percepção. Por exemplo, a sílaba pode apresentar o silêncio necessário, mas sem as transições apropriadas para um [p]. Esse tipo de configuração não promove a percepção de um [p]. Em outras palavras, o ponto importante sobre esses dados é que propriedades acústicas diferentes soam parecidas ou até mesmo iguais apenas quando a informação transmitida é igual ou similar ao que a articulação produziria.

Esses dados também mostram que um mesmo segmento acústico em diferentes contextos que especificam a atividade de articulações diferentes soa como outro som para os ouvintes. Por exemplo, a sílaba [slit] com um silêncio entre [s] e [lit] só terá um [p] percebido se houver silêncio suficiente para especificar uma oclusiva, mesmo se as transições de [p] estiverem disponíveis. No entanto, se as transições forem apresentadas isoladamente, sequer se perceberá como sons produzidos pelo trato vocal, uma vez que se ouvem frequências subindo e descendo apenas.

Fowler (1979) apresenta mais uma evidência com relação à proposta de que os ouvintes não “ouvem” pelo sinal acústico, mas sim pelo tempo da articulação. Seu argumento é sobre o experimento em que os informantes não percebiam como sincronizadas as sílabas monossilábicas alinhadas com os seus próprios *onsets* de energia acústica. No entanto, quando solicitados a sincronizar as sílabas, os informantes utilizavam as mesmas estratégias produzidas pela atividade muscular, como se utilizassem um “metrônomo imaginário”, i.e., os ouvintes parecem ouvir pelo sinal acústico o tempo dos articuladores.

Por fim, seu último exemplo em favor de a percepção seguir a articulação é o já citado *Efeito McGurk* (McDonald & McGurk, 1976). No entanto, a sua interpretação é diferente da proposta por Liberman & Mattingly (1985). Para ela, o ponto de articulação das consoantes é largamente adquirido pela visão e as vogais pela audição. Esse fato mostra ainda que é a partir da percepção da articulação que se inicia o processo de audição, não importando se a informação articulatória está presente na matriz ótica ou no sinal acústico.

3.3.2.2 *As unidades linguísticas não são literalmente articuladas*

Este ponto é uma barreira, pois uma teoria de percepção de eventos da fala não é confirmada se os constituintes da comunicação não realizarem aparições públicas. Um dos principais motivos para se acreditar que eles não aparecem está na já citada analogia da esteira de ovos de Páscoa de Hockett (1960), na qual a fala destrói a articulação livre de contexto dos segmentos quando os coarticula e sobrepõe. Assim, haveria poucas informações disponíveis para recuperar as articulações originais, livres de contexto, a não ser se levar em conta a sua mediação por processos mentais. Essa assunção é corrente na literatura e de fato se acredita que as unidades linguísticas sejam mentais. Por exemplo (Fowler, 1986: p.9):

“[Segmentos fonéticos] são *abstrações*. Eles são o resultado final de processos perceptuais e cognitivos complexos no cérebro dos ouvintes. [...] Elas [as categorias fonéticas] não têm propriedades físicas (Repp, 1981, p.1562-1463)

Os segmentos não podem ser objetivamente observados para existirem no sinal da fala ou no fluxo dos movimentos articulatórios ... O conceito de segmento é usado para suportar a priori o estudo dos aspectos físicos e fisiológicos da linguagem. [...] O segmento é gerado internamente, a criação de algum tipo perceptual – processo cognitivo. (Hammarberg, 1976: p.355)¹³⁰

¹³⁰ “[Phonetic segments] are *abstractions*. They are the end result of complex perceptual and cognitive processes in the listeners brain. (Repp, 1981, p. 1462)

They [phonetic categories] have no physical properties. (Repp, 1981, p.1463)

Segments cannot be objectively observed to exist in the speech signal nor in the flow of articulatory movements ... [T]he concept of segment is brought to bear *a priori* on the study of physical-physiological aspects of language. (Hammarberg, 1976, p. 355)

[T]he segment is internally generated, the creature of some kind of perceptual-cognitive process.

Diante, então, dessas abordagens mentalistas sobre a linguagem, que modificações teriam que ser feitas para se adotar uma abordagem realista e direta? Primeiramente, é preciso assumir que os segmentos fonéticos apenas têm propriedades que permitem que eles sejam realizados na articulação e, diante disso, a sua primeira aparição pública é como uma atividade no trato vocal. Segundo, se a produção da fala envolve a tradução de um meio mental para o domínio físico, o domínio não mental do trato vocal deve, então, ser descartado.

Com relação à primeira modificação, é preciso considerar que, como as atividades do trato vocal não são estáticas (devido a uma incapacidade do próprio trato vocal), os segmentos fonéticos não podem ser inerentemente estáticos dentro de uma teoria de eventos da fala. Por exemplo, Fowler apresenta uma imagem disponível em Carney & Moll (1971) para argumentar contra a metáfora de Hockett. Na ilustração 20¹³¹ é possível notar que há uma coerência no movimento dos articuladores, de modo que não há a destruição dos segmentos, i.e. “o conjunto de gestos para segmentos diferentes se sobrepõe, mas a separação e o ordenamento dos segmentos é preservada” (Fowler, 1986: p.10).

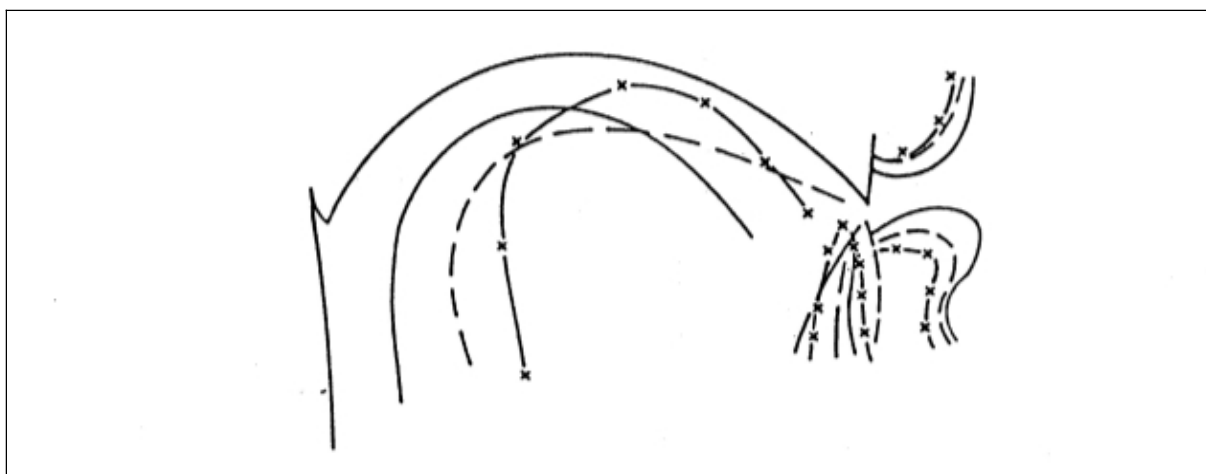


Ilustração 20: Traçado cinefluorográfico do traço vocal enquanto produzia /husi/ (redesenhado de Carney & Moll, 1971): ____ é [u]; ----, [s]/ x-x-x-x, [i] (Fowler, 1986: p.10).

(Hammarberg, 1976, p. 355)” (Fowler, 1986: p. 9).

¹³¹ Nessa ilustração apresenta-se um traçado obtido através da cinefluografia. Trata-se de uma técnica de captação dos movimentos de organismos (ou órgãos) fazendo o uso de raio-x e de uma tela fluorescente.

Com relação à segunda modificação, Fowler cita a maneira como Ryle (1949 *apud* Fowler, 1986: p.11) esquematizou a relação entre os aspectos mentais e físicos. Na sua proposta, evita-se o problema de se identificarem os processos mentais escondidos dentro da cabeça:

“Quando descrevemos pessoas a partir de qualidades mentais em exercício, não estamos nos referindo a episódios ocultos dos quais seus atos e sentenças sejam efeitos, estamos nos referindo claramente aos atos e às sentenças por si só. (Ryle, 1949: p.25)

Quando uma pessoa fala em voz alta, dá nós, simula ou esculpe, as ações que nós testemunhamos são as coisas que ele está fazendo inteligentemente ... Ele está corporal e mentalmente ativo, mas ele não está sincronizadamente ativo nos dois “lugares”, ou com dois “mecanismos” diferentes. Há uma atividade, mas ele é suscetível e requer mais que um tipo de descrição explicatória” (Ryle, 1949: p.50-51 *apud* Fowler, 1986: p.11)¹³².

Em outras palavras, é possível que haja apenas uma única atividade, mas que requer mais de um tipo de explicação. De tal modo que não se elimina a ideia de que as sentenças têm que ser planejadas, mas se elimina a ideia de que os processos encobertos são privilegiados por serem mentais, enquanto as ações que são evidentes não são.

Assim, a modificação a ser realizada para se contemplar uma teoria de percepção realista e direta é a adoção de um novo modelo de produção. O modelo de produção que se encaixa satisfatoriamente na proposta é a de Saltzman & Kelso (1983) no qual se considera a fala como a produção de um conjunto sobreposto de gestos coordenados. Nas seções seguintes (seção 3.4 e 3.6), será apresentada com maiores detalhes a proposta de produção de fala adotada pela teoria. Neste momento, só é necessário entender que Fowler rompe com a tradição ao assumir os aspectos coarticulatórios como constitutivos da produção da fala.

¹³² “When we describe people as exercising qualities of mind, we are not referring to occult episodes of which their overt acts and utterances are effects, we are referring to those overt acts and utterances themselves. (Ryle, 1949: p.25)

When a person talks sense aloud, ties knots, feints or sculpts, the actions which we witness are themselves the things which he is intelligently doing ... He is bodily active and mentally active, but he is not being synchronously active in two different 'places', or with two different 'engines'. There is one activity, but it is susceptible of and requiring more than one kind of explanatory description. (pp.51-51)” (Fowler, 1986: p.11).

3.3.2.3 *O sinal acústico não especifica os segmentos fonéticos*

Deixando de lado as questões sobre a possível destruição dos segmentos fonéticos pela articulação no momento de produção da fala, permanece o problema de que o sinal acústico parece não refletir a estrutura fonética segmental pois, por mais que os movimentos do trato vocal promovam perturbações no ar, isso não significa que tais perturbações especifiquem as causas, i.e., o movimento dos articuladores.

Na Ilustração 21, há uma esquematização desse problema. Na imagem (a), é possível notar a segmentação realizada no espectrograma ao se produzir as palavras “Santa Claus”. Como se pode notar, nessa segmentação não são considerados apenas os segmentos fonéticos; consideram-se os segmentos acústicos, i.e., suas transições e porções mediais. Com isso, em (b), notamos que os 9 segmentos fonéticos, em realidade, são especificados por 18 eventos acústicos. Nessa segunda imagem, conseguimos notar que a relação entre os segmentos fonéticos e os segmentos não é simples, pois há sobreposições e é possível que os segmentos acústicos reflitam as propriedades de mais de um segmento fonético.

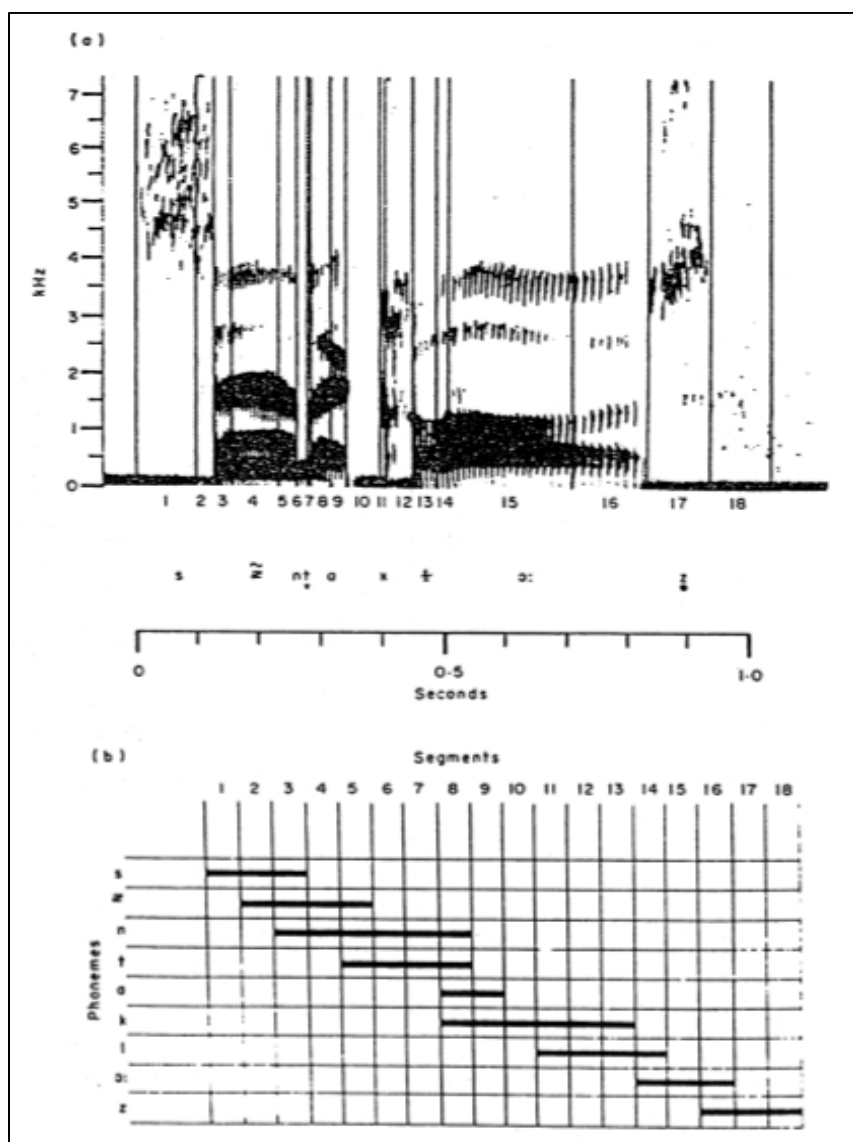


Ilustração 21: (a) Espectrograma com a produção de “Santa Claus”. (b) Visualização esquemática da relação entre os segmentos acústicos e fonéticos (reimpresso de Cutting & Pisoni, 1978 e Fant & Lindblom, 1961 apud Fowler, 1986: p.12)

Diante disso, fica a pergunta: como os ouvintes recuperam a estrutura fonética disponível no sinal acústico? Para a abordagem direto-realista, a resposta é a de que no *parsing* funcional do sinal acústico não se encontram os segmentos acústicos. A sua proposta é a de que “o *parsing* percebido deve estar no sinal; o papel especial do sistema perceptual não é criá-lo, mas apenas selecioná-lo” (Fowler, 1986: p.13).

Fowler argumenta que é possível que os segmentos fonéticos sejam recuperados no sinal acústico. Ora, não é porque há um desencontro entre essa caracterização do sinal acústico com os segmentos fonéticos que não seria possível recuperar os segmentos fonéticos. A autora cita o trabalho de Kewly-Port (1983 *apud* Fowler, 1986: p.13) que mostra uma “outra segmentação”, i.e., a partir da extração do espectro do *burst* de uma oclusiva e de seus momentos sucessivos, foi possível identificar informações sobre a continuidade da produção da consoante para a vogal. Para Fowler, há um desencontro ainda sobre como os pesquisadores devem extrair essas informações e a maneira com que os usuários da língua realmente fazem pois, para os percebedores, o sinal acústico é transparente para o acesso aos segmentos fonéticos.

Assim, para testar a hipótese de trabalho é preciso antes de tudo desenvolver dois tipos de pesquisa: (i) descrever as articulações e suas consequências acústicas; e (ii) examinar o *parsing* do sinal acústico que os ouvintes detectam. Para a autora, a pesquisa que mais se aproxima desse programa é a desenvolvida por Stevens & Blumstein (1978), na qual os autores estabelecem a caracterização de segmentos fonéticos a partir da teoria acústica de produção da fala (Fant, 1960) e desenvolvem hipóteses com relação às consequências acústicas invariantes das propriedades articulatórias essenciais dos segmentos. No entanto, essa proposta não se encaixa com aos propósitos da abordagem por eventos, pois as caracterizações sugeridas pelos autores diferem significativamente da de Fowler, uma vez que tais caracterizações são estabelecidas a partir de feixes de traços distintivos, i.e., de saída, a proposta leva em conta primitivos estáticos.

Por exemplo, Lahiri, Gwirth & Blumstein (1984 *apud* Fowler, 1986: p.14) tentam estabelecer novos invariantes acústicos a partir de um olhar dinâmico sobre o sinal acústico. Os autores examinaram a informação do espectro no tempo (“*running spectra*”) e verificaram que o ponto de articulação pode ser identificado a partir de saltos de energia de frequências altas e baixas que aparecem desde o *burst* até o início de seu vozeamento. Os próprios autores sugerem que esses achados ilustram que os invariantes são dinâmicos, pois suas características se distribuem no tempo e variam de acordo com vizinhança da soltura da oclusiva. Esse estudo, para Fowler, mostra que os resultados de Stevens & Blumstein (1978)

podem ser distintos se eles tivessem partido de um olhar dinâmico sobre os dados.

3.3.2.4 *A percepção envolve processos top down* e os percebedores realizam erros

Fowler salienta que os processos *top down* são aqueles que fazem inferências sobre os fonemas que não são produzidos e fazem correções sobre as palavras produzidas de maneira errada. Diante do fato de que o que não é produzido não é percebido dentro da perspectiva realista, tais dados são um verdadeiro desafio para a abordagem proposta por ela. No entanto, a autora argumenta que ainda é preciso estudar mais os processos *bottom-up* para que se entenda como são preenchidos os erros ou os segmentos não pronunciados quando um processo *top down* não influencia uma produção.

Citando Shaw, Turvey & Mace (1982 *apud* Fowler, 1986: p.15), a autora argumenta que sua proposta se preocupa mais com as distinções entre o que os percebedores podem fazer e o que eles devem fazer, ou seja, “há a necessidade de um suporte informacional para a atividade ser capaz de ser diretamente extraída do meio informacional e para a percepção ser nada mais que a extração direta de informação da estimulação proximal” (Fowler, 1986: p.15). Desta forma, os efeitos *top down* sobre os erros de produção são, nessa perspectiva realista, uma atividade dos percebedores, uma vez que, em ambientes conhecidos, eles baseiam suas atividades sobre o que o ambiente normalmente proporciona, i.e., eles conhecem o que o ambiente oferece e, por isso, sabem como preencher o que falta ou como corrigir um dado erro.

Como se pode notar, essa barreira não é facilmente superada. Entretanto, a autora sugere que a sua superação depende de mudanças no olhar sobre os processos da língua. Em suas palavras:

“É importante enfatizar que a tentativa precedente de superar a quarta barreira pretendeu fazer mais do que traduzir uma descrição de processos *top down* e *bottom-up* em uma terminologia mais palatável aos teóricos do evento. Além disso, tento conceder um papel de informação que não está na estimulação para orientar a atividade, enquanto preservo as ideias de que a

percepção por si só deve ser direta e por isso, sem esforço, e que a atividade pode ser (mas frequentemente não é) guiada exclusivamente pelas disposições percebidas” (Fowler, 1986: p.16)¹³³.

3.3.3 Como a percepção guia a ação?

Ao contrário das barreiras apresentadas nas seções anteriores, os argumentos sobre como a percepção guia a ação dos usuários de uma língua são mais facilmente depreendidos em uma abordagem de eventos reais.

Primeiro, uma maneira de se ter a ação guiada pela percepção é controlando os seus próprios articuladores da mesma maneira que o seu interlocutor. Para Fowler, o princípio que guia esse aspecto da proposta é a imitação, uma vez que pesquisas mostram que os falantes, durante um ato comunicativo, passam a se comportar da mesma maneira que o seu interlocutor. Tais ajustes são chamados de sincronia interacional e eles ocorrem com relação à velocidade de fala, ao volume e à utilização e duração das pausas.

Segundo, a autora argumenta que há uma predisposição em realizar imitações desde a infância. O trabalho de MacKain, Studdert-Kennedy, Spieker & Stern (1983 *apud* Fowler, 1986: p.21) mostra que crianças estabelecem relações entre o som ouvido e a articulação quando submetidos a um som e duas imagens diferentes. Elas tendem a olhar a imagem que está articulando o som que está sendo produzido. Ainda mais, o trabalho de Meltzoff & Moore (1985 *apud* Fowler, 1986: p.21) revela que as crianças têm conhecimento sobre os seus próprios articuladores, pois elas imitam gestos faciais, tais como a protrusão de lábios e língua, mesmo sem poder ver seus próprios gestos¹³⁴. De acordo com Fowler, “esses achados sugerem que as crianças deveriam ser capazes de imitação vocal” (Fowler, 1986: p.21).

¹³³ “It may be important to emphasize that the foregoing attempt to surmount the fourth barrier is intended to do more than translate a description of top-down and bottom-up processes into a terminology more palatable to event theorists. In addition, I am attempting to allow a role for information not currently in stimulation to guide activity, while preserving the ideas that perception itself must be direct and hence, errorless, and that activity *can be* (but often is not) guided exclusively by perceived affordances” (Fowler, 1986: p.16).

¹³⁴ Um exemplo disso é o fato de as crianças mandarem beijos ou mostrarem a língua ao adulto que as manda beijos ou as mostra a língua.

Terceiro, Fowler apresenta como argumento os fatos achados de Labov (1963 *apud* Fowler, 1986: p.22) sobre o fato de que o estilo de uma pessoa pode variar de acordo com as respostas dadas em um contexto conversacional, podendo ser guiado pelo estilo do seu companheiro na conversação. Isso ocorre pois, “em um cenário conversacional natural, os falantes usam a variação fonética com fins psicológicos e sociais e, necessariamente, dado que os ouvintes são sensíveis a esses usos” (Fowler, 1986: p.22).

Por fim, a autora levanta duas questões para finalizar a sua proposta. São elas: (i) o que permite à atividade foneticamente estruturada do trato vocal servir como mensagem significativa?; e (ii) uma mensagem linguística da fala pode ser diretamente percebida?

Para responder à primeira questão, Fowler sugere (a partir das propostas de Gibson, 1966) que é necessário antes de tudo estabelecer uma relação entre a estrutura na mídia informacional com as propriedades dos eventos. Essa relação é satisfeita ao relacionar as atividades do trato vocal com as suas consequências acústicas. Embora essas relações ainda não estejam traçadas (pois este é um dos pontos a serem desenvolvidos pela teoria), ela argumenta que se trata de uma via passível de desenvolvimento: a atividade do trato vocal pode servir como mensagem fonológica (ou linguística) se consideramos que a competência linguística dos falantes e ouvintes de uma língua é a capacidade de controlar seus tratos vocais, i.e., eles produzem gestos com significância linguística. Outra possibilidade está justamente na referência ao nicho ecológico que suporta a comunicação:

“A atividade do trato vocal apenas pode constituir uma mensagem linguística em um cenário no qual, historicamente, a atividade do trato vocal apropriadamente restringida fez trabalho linguístico. A habilidade de um ouvinte extrair a mensagem linguística da atividade do trato vocal pode ser entendida como uma abordagem 'conceitual' (eu diria 'funcional') pelas linhas seguintes: ouvintes apreendem o trabalho linguístico que a atividade foneticamente estruturada do trato vocal está fazendo em virtude de suas sensibilidades ao contexto social e histórico das restrições nas quais a atividade é realizada” (Fowler, 1986: p.23)¹³⁵.

¹³⁵ “Vocal-tract activity can only constitute a linguistic message in a setting in which, historically, appropriately constrained vocal-tract activity has done linguistic work. A listener's ability to extract a linguistic message from vocal-tract activity may be given a 'conceptual' (I would say 'functional') account along lines such as the following: listeners apprehend the linguistic work that the

A resposta da segunda pergunta cai em uma discussão mais ampla que disserta se percebemos o mundo direta ou indiretamente. A autora tenta minimizar as diferenças entre as duas concepções, expondo a proposta de Gibson (1966 *apud* Fowler, 1986: p.24) de que a percepção é indireta (não no sentido de ser necessário um processo cognitivo para fazer a mediação), mas sim em perceber a informação sobre os eventos a partir de uma mídia informacional que apresenta as estruturas de uma fonte distal. Em suas palavras:

“Quando um evento é percebido diretamente, ele é percebido pela extração de informação para o evento a partir de uma mídia informacional. Quando a comunicação linguística é indiretamente percebida, a informação sobre as atividades do trato vocal do falante são extraídas a partir do sinal acústico. O trato vocal (por hipótese) constitui palavras foneticamente estruturadas em sequências gramaticais, e assim especifica indiretamente o quer que seja sobre a sentença” (Fowler, 1986: p.24)¹³⁶

Para ela, o que é digno de ênfase é a relação entre a sentença e o que ela significa, de modo que o significado não tem que estar adequado às leis físicas diretamente, mas sim “às restrições culturais que evoluíram sobre as gerações de uso da linguagem” (Fowler, 1986: p.24) que devem estar disponíveis a qualquer usuário da língua.

Em suma, a distinção entre o direto e o indireto pode ser dissolvida se pensarmos que nas duas abordagens a atividade foneticamente estruturada do trato vocal e a informação linguística são diretamente percebidas pela extração da informação invariante a partir do sinal acústico, embora as origens das informações sejam distintas: a estrutura fonética é proveniente das relações coordenadas entre os articuladores e a mensagem linguística pelas restrições que refletem o contexto cultural moldado pelo uso da linguagem através das gerações. Por exemplo, se alguém realiza um enunciado sobre uma mesa, de fato não apreendemos a mesa diretamente, pois não se experiencia a mesa. O que há neste caso é a percepção da

phonetically structured vocal-tract activity is doing by virtue of their sensitivity to the historical and social context of constraint in which the activity is performed” (Fowler, 1986: p. 23).

¹³⁶ “When an event is perceived directly, it is perceived by extraction of information for the event from informational media. When a linguistic communication is indirectly perceived, information for the talker’s vocal-tract activities is extracted from an acoustic signal. The vocal-tract activity (by hypothesis) *constitutes* phonetically structured words organized into grammatical sequences, and thereby indirectly specifies whatever the utterance is about” (Fowler, 1986: p. 24).

perspectiva do falante. Diante disso, ela sugere que nada é percebido indiretamente.

Nesta seção (e nas anteriores), há a primeira proposta mais sistematizada sobre a Teoria do Realismo Direto da Percepção da Fala. Como se pode notar, há uma tentativa de se justificar e apresentar um programa para o desenvolvimento da teoria. A questão recorrente nos argumentos é a falta de estudos sobre as atividades do trato vocal na produção da fala e as suas consequências acústicas. Para Fowler, ao entender como se organizam as atividades do trato vocal e ao identificar quais são as suas consequências acústicas, a sua teoria de percepção ganha mais força.

É importante mencionar que a sua abordagem aponta para o desenvolvimento de uma teoria fonológica que leve em conta os aspectos dinâmicos da produção da fala e esse modelo é a Fonologia Articulatória de Browman & Goldstein (1986 e seguintes). Passaremos à explanação desse modelo na seção seguinte (3.4 do Capítulo 2).

3.4 Fonologia Articulatória

A Fonologia Articulatória (Browman & Goldstein 1986, 1989, 1990, 1992) tem sido formulada e apresentada em textos desde meados da década de 1980. Nesse período, ela sofreu reformulações e foi em 1992 que ganhou maior visibilidade e “autonomia”, pois é nesse texto que os autores vão sistematizar sua proposta e, com isso, delimitar a que veio a teoria. Apresentaremos, então, a sua sistematização de 1992 que versa praticamente sobre a produção da fala. Somente em textos seguintes (como o de Goldstein & Fowler – 2003) que o gesto articulatório será tratado como unidade de produção e de percepção¹³⁷.

Seu desenvolvimento ocorre pela constatação de que há fatos nas línguas do mundo que não são representados de maneira parcimoniosa quando se assumem primitivos não dotados de tempo intrínseco¹³⁸. Assim, esse passa a ser o ponto

¹³⁷ Antes de apresentarmos o trabalho de Goldstein & Fowler (2003), na seção 3.6 (do Capítulo 2), trataremos do conceito de paridade (Lieberman & Whalen, 2000) na seção 3.5 (do Capítulo 2), pois trata-se de uma noção que justifica a proposição de uma unidade comum à produção e à percepção da fala.

¹³⁸ O trabalho que discute detalhadamente as vantagens de se adotar primitivos de análise fonológica dotados de tempo intrínseco se encontra em Fowler (1980).

principal dos textos das primeiras versões¹³⁹ do modelo. Tendo isso em vista, os autores tentam justificar a adoção de um novo primitivo de análise e também mostrar que os gestos podem funcionar como unidades discretas que organizam a fonologia das línguas. Vale salientar que, embora seja inovadora a proposta, os autores ainda permaneciam presos a uma tradição herdeira de uma fonologia de traços, pois a todo momento argumentam sobre a proximidade que há entre um tratamento dentro de uma fonologia de arcabouço gestual com uma fonologia de traços não linear (como a Geometria de Traços de Sagey, 1986¹⁴⁰). A argumentação desse período inicial chega à proposta, no mínimo curiosa, de uma Geometria do Tubo - ideia esta abandonada sem justificativas nos textos seguintes ao de 1986.

Um outro ponto (herdado de uma Geometria de Traços) abandonado sem justificativas é a noção de hierarquização dos gestos, i.e., os gestos orais estariam em posição mais alta da hierarquia, pois eles seriam os principais responsáveis pela organização rítmica da fala. Por sua vez, os gestos vélicos estariam em posição mais baixa da hierarquia, por não contribuírem massivamente para a organização do ritmo da fala¹⁴¹.

Deixando de lado as discussões do início da formulação do modelo, verificamos que é em 1992 que a Fonologia Articulatória (doravante FAR) ganha formalização sistematizada e é nesse momento que ela “rompe” definitivamente com a tradição Gerativa, principalmente em dois pontos: 1) o modelo passa a assumir que não há a necessidade de se considerar a existência de um módulo fonológico, pois apenas a existência de um único módulo fonético¹⁴² - que implementaria diretamente as pautas gestuais - seria suficiente para representar a fala; 2) e, como a representação da fala teria como primitivo de análise o gesto articulatório, o

¹³⁹ Utilizo aqui “versões” da teoria devido ao próprio percurso de formulação da teoria. Para um histórico do percurso do modelo ver Silva (2002).

¹⁴⁰ Para uma revisão em língua portuguesa sobre esse modelo e seus limites, consultar D’Angelis (1998).

¹⁴¹ Esse ponto já foi mencionado anteriormente na seção 3.1. do Capítulo 2 (nota 91), onde se comenta a assunção tácita da FAR de considerar os gestos vocálicos como organizadores da cadeia da fala. Cabe mencionar aqui que isso não é apenas um “herança” de uma Geometria de Traços, mas sim um ponto que converge para as mesmas propostas de Fowler (1986), mostrando a “sintonia” com que o modelo de produção (a FAR) e de percepção (a TRD) se desenvolvem.

¹⁴² Em 1992, Browman & Goldstein propunham a existência de um único módulo fonético. A introdução da noção da existência de um módulo fônico (isto é, fonético e fonológico) se encontra em Albano 2001. Assim, quando me refiro neste trabalho à proposta de um nível fonético proposto por Browman & Goldstein, estou considerando esse período do desenvolvimento da teoria.

modelo conseguiria dar conta satisfatoriamente das alofonias contínuas sem custos de processamento, pois a arquitetura do modelo é implementacional e não procedural - de modo que não há um conjunto de regras a serem derivadas para se chegar a um *output*. Tais afirmações são corroboradas pelos dados sobre as alofonias contínuas envolvidas na produção das variantes *dark* e *light* da lateral do inglês (Sproat & Fujimura, 1993) e pelas diferenças entre as consoantes nasais de *onset* e de coda do inglês (Krakow, 1989 *apud* Browman & Goldstein, 1992: p.167).

Por mais que a “natureza” do primitivo de análise assumido pela teoria tenha sido apresentada “pouco a pouco” nos trabalhos de 1986, 1989, 1990 e 1992, há aqui um outro ponto de divergência importante das teorias herdeiras de um arcabouço gerativo, pois as alterações sugeridas pelos autores quanto à produção da fala só são possíveis se o gesto articulatório for tomado como primitivo de análise: pois “são também caracterizações abstratas de eventos articulatórios, dotados de tempo intrínseco ou duração” (Browman & Goldstein, 1992: p.155). Ou seja, “os gestos são eventos que não são encontrados na produção da fala e as suas consequências podem ser observadas nos movimentos dos articuladores envolvidos nela. Esses eventos consistem na formação e soltura de constrições no trato vocal” (Browman & Goldstein, 1992: p.156).

Com isso, os gestos articulatórios teriam seus movimentos definidos através da Dinâmica de Tarefa (Saltzman, 1986, Saltzman & Kelso, 1987; Saltzman & Munhall, 1989 *apud* Browman & Goldstein, 1992: p.156), pois trata-se de um modelo de base psicológica que consegue representar diferentes tipos de ações que envolvem a coordenação de vários articuladores que têm por objetivo realizar uma tarefa. No caso da fala, são as tarefas envolvidas na formação das constrições relevantes para cada língua que interessam para a teoria. Com isso, o gesto articulatório é modelado a partir de uma equação dinâmica que descreve seus movimentos a partir de um sistema massa-mola que, ao ter uma massa presa a uma extremidade da mola e a outra a uma superfície qualquer, descreve um movimento oscilatório ao se puxar e, em seguida, soltar a mola. As diferentes trajetórias produzidas pelo sistema são possíveis com a alteração nos valores relativos à massa do objeto, rigidez da mola, comprimento da mola na posição de repouso e amortecimento do sistema:

$$m\ddot{x} + b\dot{x} + k(x - x_0) = 0$$

sendo: m = massa do objeto;

b = amortecimento do sistema;

k = rigidez da mola;

x_0 = comprimento da mola na posição de repouso;

x = deslocamento instantâneo do objeto;

\dot{x} = velocidade instantânea do objeto;

\ddot{x} = aceleração instantânea do objeto.

É a partir da definição da equação dinâmica que os autores acabam reconhecendo o movimento de variáveis do trato e não o movimento de articuladores individuais. Assim, “a variável do trato caracteriza a dimensão da constrição do trato vocal; os articuladores que contribuem para a formação e soltura dessas constrições são organizados em estruturas coordenadas” (Browman & Goldstein, 1992: p.156). Dessa forma, a variável do trato “abertura labial”, por exemplo, é afetada pela ação de três articuladores: lábio superior, lábio inferior e mandíbula. Assim, os autores acabam estipulando as variáveis do trato e os articuladores responsáveis por implementá-las na produção da fala, conforme disposto na tabela 2:

Variável do trato	Articuladores envolvidos
PL – protrusão labial	Lábios inferior e superior, mandíbula
AL – abertura labial	Lábios inferior e superior, mandíbula
LCPL – Local de constrição de ponta de língua	Ponta e corpo de língua, mandíbula
GCPL – grau de constrição de ponta de língua	Ponta e corpo de língua, mandíbula
LCCL – local de constrição de corpo de língua	Corpo de língua, mandíbula
GCCL – grau constrição de corpo de língua	Corpo de língua, mandíbula
VEL – abertura de véu palatino	Véu palatino
GLO – abertura de glote	Glote

Tabela 2: Variáveis do trato e articuladores envolvidos na realização da tarefa especificada pelas variáveis¹⁴³.

¹⁴³ As variáveis do trato apresentadas nos trabalhos de Browman & Goldstein (1986 e seguintes) fazem parte do programa de síntese articulatória desenvolvido pelo grupo dos Laboratórios Haskins, o TADA (*Task Dynamic Application*). Para maiores informações, consultar: http://www.haskins.yale.edu/tada_download/index.php. Nas páginas que tratam da proposta da FAR (seção 3.4 do Capítulo 2), seguiremos a utilização terminológica utilizada pelos próprios autores que, ao invés de se valerem das siglas sugeridas em seu modelo, acabam utilizando tanto as variáveis do trato referentes ao local e grau de construção como gestos. Nesse caso, as variáveis do trato LCPL e GCPL são substituídas por PL (Ponta de Língua) e LCCL e GCCL por CL (Corpo de Língua). Os graus de constrição são especificados nas pautas gestuais pelos

O gesto articulatório também ganha em sua formalização descritores gestuais que lhe seriam atribuídos, de modo que é possível estabelecer contrastes entre os gestos. Dessa forma, são estabelecidos os descritores gestuais [fechado], [crítico], [estreito], [médio] e [largo] para grau de constrição e [protruso], [labial] e [dental] para os lábios; [labial], [dental], [alveolar], [pós-alveolar] e palatal para ponta de língua; e [palatal], [velar], [uvular] e [faríngeo] para corpo de língua.

Assim, para Browman & Goldstein, tendo definido o gesto articulatório a partir das tarefas a serem realizadas nas variáveis do trato, se torna possível concebê-lo como unidade contrastiva de itens lexicais, de modo eles diferem pela: a) presença ou ausência de um gesto; b) pelos diferentes valores das variáveis da equação dinâmica; e c) pelas diferentes organizações temporais possíveis entre os gestos - o faseamento. Vale salientar que tais parâmetros de diferenciação entre os itens lexicais somente são possíveis devido à incorporação do tempo intrínseco ao primitivo de análise. Steriade (1990) argumenta que essa é a principal diferença (e ganho) da Fonologia Articulatória (FAR) com relação às outras teorias, pois os outros modelos (que adotam o traço como primitivo) acabam precisando de uma camada temporal extrínseca aos seus primitivos para dar conta de consoantes geminadas, por exemplo.

Dessa forma, os gestos articulatórios são formalizados em pautas gestuais (na sequência), que consistem na representação do tempo de ativação dos gestos no eixo horizontal, e no eixo vertical, da magnitude dos gestos, ou seja, é possível contemplar na representação o tempo de ativação de cada gesto, além de uma eventual sobreposição temporal entre gestos.

Com isso, podemos notar, na Ilustração 22, que temos a produção de um gesto de corpo de língua (CL) largo e faríngeo correspondente à vogal da palavra “bad” do inglês e ao mesmo tempo que se inicia o gesto de CL, há a ativação do gesto labial fechado (início da caixinha no sentido horizontal da pauta), que acaba caracterizando a oclusão de [b] da palavra. Quase ao fim da palavra há a ativação do gesto de ponta de língua (PL) referente ao [d] e também podemos constatar que ele se estende até o fim do gesto de CL. É interessante notar que, ao contrário de especificadores que são descritos dentro das “caixinhas”.

outros modelos de análise fonológica, aqui é possível contemplar os fenômenos que, em princípio, não são contemplados por outras, a saber: as co-produções dos gestos de lábios, de CL e de PL. Dessa forma, se dá conta tanto das características dinâmicas da fala como dos fatos categóricos que as noções de fonema ou matrizes de traços sugerem.

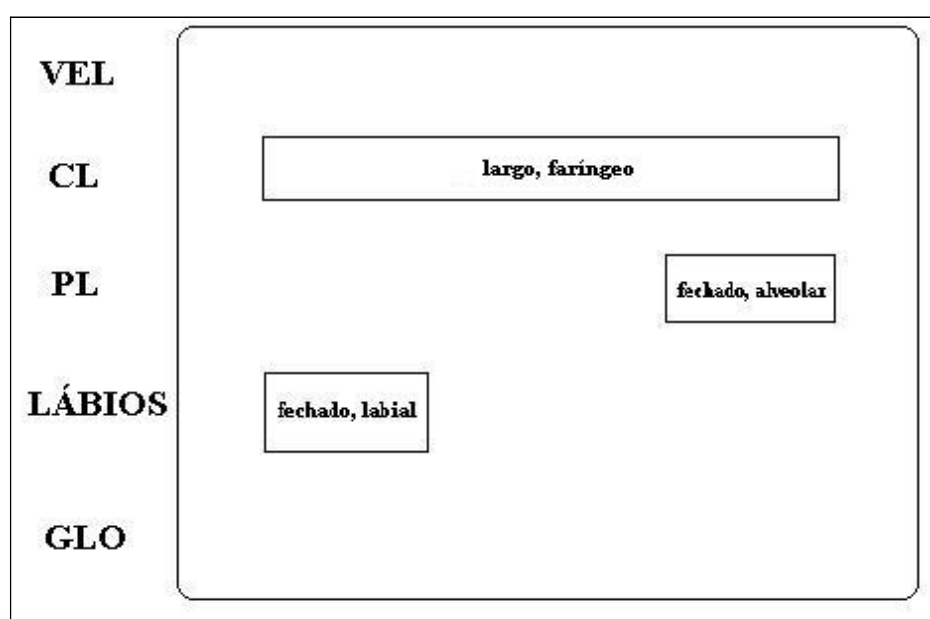


Ilustração 22: Pauta gestual da palavra "bad" (baseada em Browman & Goldstein, 1992: p.158).

Já a pauta gestual da Ilustração 23 mostra que é possível contemplar contrastes lexicais no modelo. Podemos verificar que a pauta da palavra "pad" do inglês apenas se diferencia da pauta de "bad" pela ativação do gesto longo de glote (GLO), que produz a oclusiva surda bilabial [p]. Podemos notar também que o gesto de GLO inicia depois do início dos gestos de CL e PL, evidenciando que um segmento surdo não se caracteriza por seu ensurdecimento completo e tampouco acaba com o fim da oclusão bilabial. Neste caso específico do inglês, a maneira como o gesto de GLO está faseado ainda promove a aspiração de [p]. Caso estivéssemos diante de uma língua que não realiza a aspiração das oclusivas surdas, GLO receberia outro tipo de faseamento, i.e., o gesto terminaria antes ou ao mesmo tempo do fim do gesto de LÁBIOS.

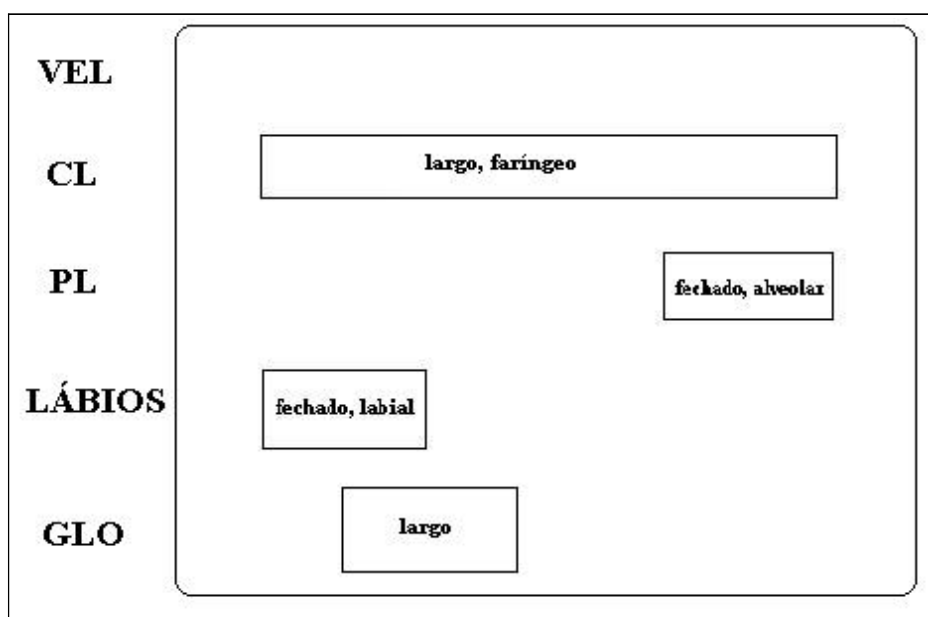


Ilustração 23: Pauta gestual da palavra “pad” (baseada em Browman & Goldstein, 1992: p.158).

É tendo isso em vista que Browman & Goldstein acabam propondo que há apenas um nível fonético nas gramáticas das línguas, pois se o gesto é uma unidade contrastiva e pré-linguística¹⁴⁴, é possível que as crianças consigam fazer emergir os contrastes necessários para se produzir a fala em determinada língua. Para os autores, a aquisição da fala acompanharia a maturação neuromotora da criança, contribuindo para que a linguagem emergisse do fonético para o abstrato e, como o gesto articulatório possui tanto características concretas como abstratas, tal ligação só seria possível com a adoção do gesto articulatório como primitivo de análise¹⁴⁵. Esse ponto mostra que, tanto para a FAR quanto para a TRD, a criança consegue produzir e perceber a fala porque ela “conhece” os seus articuladores. Para as duas abordagens, o fato de a criança amadurecer seu sistema motor durante a amamentação e de reconhecer nas outras pessoas a sua própria articulação (quando responde ao beijo de um pessoa, por exemplo) é de extrema importância

¹⁴⁴ Os autores consideram como unidades pré-linguísticas os gestos envolvidos em tarefas que serão utilizadas na produção da fala. Por exemplo, a sucção dos bebês no ato de mamar é considerada relevante linguisticamente no período anterior à produção de fala, pois são constitutivas do amadurecimento do sistema neuromotor da criança.

¹⁴⁵ Para maiores considerações sobre a adoção do gesto articulatório como primitivo de análise e a emergência da linguagem consultar: Studdert-kennedy & Goldstein (2003).

para que produção e percepção da fala compartilhem de uma base comum. Mas o que garantiria e qual vantagem haveria de esses dois processos compartilharem de uma mesma unidade? A resposta é dada pelo conceito de *paridade*. Na próxima seção (3.5 do Capítulo 2), trataremos dele.

3.5 Paridade: o conceito que faltava

O modelo de Fowler (1986) para a percepção da fala ainda parecia “implausível” no momento da sua proposição e muito havia que ser feito para que o seu programa de pesquisa ganhasse força e adesão na comunidade científica. No entanto, a proposta de Browman & Goldstein (1992), por sua vez, parece causar “barulho” e apontar para uma nova maneira de olhar a produção da fala. A FAR imprime nos estudos fônicos a possibilidade de superar uma das barreiras enfrentadas pela TRD, de modo que o modelo busca conhecer as características básicas dos movimentos coordenados realizados pelos órgãos do trato vocal. Para Fowler, é o conhecimento dessas características e de suas consequências acústicas que possibilitará o desenvolvimento de uma abordagem realista e direta para a percepção da fala. Assim, ao se propor um novo primitivo de análise fonológica (que implode com a distinção dicotômica entre Fonética e Fonologia), a TRD começa a “sair do papel” e a desenvolver seu programa de pesquisa.

Contudo, uma pergunta ainda permanecia: se os gestos articulatórios são as unidades de produção e de percepção da fala, como se estabelece a troca dessas unidades no ato de comunicação? A resposta inicial de Fowler (1986) era a de que os usuários da língua tinham suas ações guiadas pela percepção desses movimentos e o princípio que a guiava era a imitação.

Evidentemente, essa resposta não agradava os pesquisadores, de modo que a TRD parecia apontar para uma abordagem associacionista¹⁴⁶ da mente. Também vale ressaltar que o “simples” fato de se adotar novos primitivos sobre a produção e percepção da fala não garantia que tais primitivos fossem trocados no ato comunicativo. Ainda mais, as propostas sobre a percepção e a produção não

¹⁴⁶ Ver seção 2.3.2.1 (do capítulo 2) para a explicação de Fodor sobre as maneiras de se estruturar a mente.

pareciam se encaixar com relação a um modo de se estruturar a mente: enquanto a FAR parecia apontar para um aparente horizontalismo (pois havia abandonado a dicotomia do fonético/fonológico); a TRD se direcionava a um aparente verticalismo (porque deixava de lado as ondas sonoras como dotadas de características suficientes para a percepção da fala).

Esse ecletismo teórico da abordagem dinâmica permitia muito mais a proposição de perguntas do que de soluções. Perguntas do tipo “*como barrar uma alofonia se ela é contínua?*” ou “*se há influência do sistema fonológico de uma língua sobre a percepção de outra, como modelar esse fenômeno se se adota apenas um módulo fonético?*” não encontravam explicações convincentes. Essa falta de explicações advinha da incompletude do modelo realista e direto (pois ainda buscava superar as suas barreiras através do desenvolvimento do seu programa de pesquisa) e da “nebulosidade” da proposta da FAR, que ainda não havia sido compreendida completamente pelos pesquisadores.

No entanto, Liberman & Whalen (2000) propõem um conceito que permite a uma teoria de produção de fala como a FAR e a uma de percepção como a TRD se encaixarem satisfatoriamente. Tal conceito é o de *paridade*.

O conceito de paridade, em realidade, é um princípio que regulamenta qualquer sistema de comunicação, i.e., seria uma exigência (moldada pela evolução) que garantiria a troca de unidades de mesma natureza entre a produção e a percepção. Nas linhas a seguir, apresentaremos a sua argumentação a favor do conceito de paridade e de como ele possibilita que se unifique o tratamento da produção e percepção da fala (a ser apresentado na próxima seção – 3.6). Esse conceito passa a ser importante para a TRD porque assim há uma garantia de que não é preciso haver transformações (tal como a proposta da TM no arcabouço gerativo-transformacional) do nível acústico para o articulatório e vice-versa.

A proposta dos autores é a de que cada espécie evoluiu seu sistema comunicativo para um fim específico e, sendo a fala um dos produtos da evolução dos seres humanos, ela também estaria sujeita a essa especialização. Por exemplo, cada especialização explora um ou mais meios para transmitir os seus sinais padronizados de comunicação. “Assim, algumas criaturas usam sinais acústicos,

algumas ópticas, algumas mecânicas, algumas químicas, e o peixe elétrico o faz eletricamente” (Liberman & Whalen, 2000: p.190).

No entanto, a grande diferença entre os sistemas dos outros animais e o dos humanos é que aqueles são fechados, pois “todo sinal é significativo, e sinais nunca são rearranjados para transmitir novos significados (Liberman & Whalen, 2000: p.190). Com a linguagem humana ocorre justamente o oposto, porque a “linguagem é aberta ou gerativa, sendo capaz de, em princípio, codificar um número infinito de significados” (Liberman & Whalen, 2000: p.190).

Essa capacidade está vinculada a outra diferença primordial com relação aos sistemas humanos e não humanos. Trata-se do fato de a evolução ter separado a forma do sinal da sua função semântica (cf. Studdert-Kennedy, 2000 *apud* Liberman & Whalen, 2000: p.190), i.e., durante a evolução, os sinais foram criados com aspectos comunicativos distintos que possibilitaram à fala seguir uma estratégia combinatória. Esse princípio é muito semelhante ao utilizado para se estudar os compostos químicos e a genética.

“Portanto, na fala, poucos segmentos sem significado são combinados e permutados variadamente para formar um vocabulário de palavras com significado indefinidamente grande e expansível, que, por sua vez, promove a base necessária para a continuação da aplicação do princípio de formação de proposições através dos processos combinatórios da sintaxe” (Liberman & Whalen, 2000: p.190)¹⁴⁷.

Sendo mais específico ainda sobre a linguagem humana, é preciso salientar que o conceito de paridade surge para tentar responder às perguntas: como os falantes e ouvintes sabem o que conta no ato comunicativo?; como eles são capazes de se comunicarem em um mesmo código?; e como a produção e a percepção se desenvolveram juntas na evolução? Os autores sugerem que há três exigências para que haja paridade e elas são melhor acomodadas em uma estrutura mental verticalista¹⁴⁸.

¹⁴⁷ “Thus, in speech, a few meaningless segments are variously combined and permuted to form an indefinitely large and expandable vocabulary of meaningful words, which, in turn, provides the necessary base for the further application of the principle to the formation of propositions through the combinatorial processes of syntax” (Liberman & Whalen, 2000: p.190).

¹⁴⁸ A disputa entre verticalismo e horizontalismo não faz parte da agenda de discussões da TRD. Contudo, decidimos reproduzi-la aqui justamente para marcar o momento em que essas questões são colocadas no trabalho de Liberman & Whalen (2000), que, por sinal, são adeptos da TM. O importante desse conceito é que ele foi “incorporado” para que o sistema realista e direto funcione.

A primeira delas se aplica quando há comunicação em uma única via entre um falante e um ouvinte, ou seja, trata-se da situação em que uma pessoa fala e a outra apenas ouve. Nessa modalidade a comunicação só pode ter sucesso se as duas partes envolvidas compartilharem de um entendimento comum sobre o que conta e o que não conta, ou seja, os dois participantes têm que saber que, por exemplo, um /ba/ conta e uma fungada, não.

Para uma abordagem horizontalista, o problema está no fato de que tanto uma sílaba como /ba/ quanto uma fungada elicitam processos perceptuais auditivos, mas apenas uma delas (/ba/) apresenta relevância linguística. Para o horizontalista, a explicação de como uma sílaba (ou uma unidade sonora qualquer) passou a receber relevância linguística ou significado fonético se refere à hipótese de que um primeiro ancestral falante deliberou por consenso com outros da mesma espécie quais seriam as unidades fonéticas. Nessa perspectiva, a fala seria um artefato, tal como as letras do alfabeto.

Já para uma abordagem verticalista, “o falante produz gestos que são manejados pelo mecanismo que evoluiu especificamente em conexão com uma função comunicativa (linguística), [...], sendo os gestos fonéticos na sua própria natureza” (Lieberman & Whalen, 2000: p.189). A proposta também segue na direção de sugerir que os gestos são recuperados no sinal acústico através de um componente especializado para a linguagem. Nesta abordagem, então, não há necessidade de se apelar para conexões entre “uma representação auditiva inicial e alguma unidade fonética abstrata” (Lieberman & Whalen, 2000: p.189), uma vez que o primitivo é inicialmente fonético. Trata-se da proposta de que existe um módulo fonético específico para a fala. Aqui, ele está atrelado a um módulo linguístico maior que, ao invés da versão fodoriana, foi moldado pela evolução.

A segunda exigência para a paridade se aplica em um contexto de comunicação em que há dois participantes ativos, ou seja, falantes e ouvintes trocam de papéis em turnos de fala. Nessa modalidade, é essencial que as representações fonéticas presentes no cérebro de um dos participantes sejam replicadas no cérebro do outro, caso contrário, eles não podem se comunicar em um

Aliás, por mais que o texto de Lieberman & Whalen discutam questões caras à TM, temos nesse trabalho explicações de base evolucionista que são tranquilamente acomodadas na abordagem da TRD.

mesmo código.

Para a abordagem horizontalista, essa exigência não é acomodada satisfatoriamente, pois a representação de uma das partes é puramente motora (implementação da fala), enquanto a representação da outra é auditiva (recepção das ondas acústicas). Uma possível resposta horizontalista para a questão é a de que esses fenômenos estão atrelados às unidades fonéticas. Entretanto, como foi possível notar na primeira exigência, não há explicação convincente com relação à origem e à conexão das unidades com as representações motoras e auditivas.

Alternativamente, a abordagem verticalista permite ver que as duas partes envolvidas no ato de comunicação negociam com uma mesma moeda de troca: os gestos fonéticos. Assim,

“A visão vertical, alternativamente, nos permite ver que as partes conduzem seu negócio na moeda comum dos gestos fonéticos: o gesto, que é a unidade de produção para o falante, é a unidade de percepção para o ouvinte. Não é necessário para nenhum deles associar representações motoras e auditivas díspares a unidades fonéticas com as quais eles compartilham uma representação que é, sob qualquer ponto de vista, completamente arbitrária” (Liberman & Whalen, 2000: p.190)¹⁴⁹.

Por fim, a terceira exigência diz respeito ao processo de coevolução requerido no desenvolvimento biológico da fala. Isso significa que a produção e a percepção da fala devem ter evoluído “de mãos dadas”. Caso contrário, o sistema teria chegado a um impasse.

Diante disso, não há saída para os horizontalistas, pois, ao assumirem representações distintas para os aspectos motores e perceptuais, eles aceitam o impasse de que os dois processos (produção e percepção da fala) não compartilham de primitivos de mesma natureza e que produção e percepção podem ter evoluído separadamente. As suas explicações são as que os falantes e ouvintes estabelecem conexões a partir da experiência das consequências auditivas das suas próprias manobras articulatórias. Trata-se de uma abordagem de aprendizado por associação

¹⁴⁹ “The vertical view, alternatively, permit us to see that the parties conduct their business in the common currency of phonetic gestures: the gesture that is the unit of production for the speaker is the unit of perception for the listener. There is no need for either to refer grossly dissimilar motor and auditory representations to phonetic units to which they bear a relationship that is, from any point to point, wholly arbitrary” (Liberman & Whalen, 2000: p. 190).

que não se encaixa com uma abordagem evolucionista (tal como propõe a TRD a partir de Gibson – 1966).

Assim, o fato de produção e percepção terem coevoluído é facilmente acomodado em uma abordagem verticalista, porque, ao se assumir que a representação primária do falante e do ouvinte é exatamente a mesma (o gesto), a troca de informações a partir de uma das partes é inevitavelmente espelhada pela outra. Em suma, a paridade se apresenta como uma exigência evolucionista para que produção e percepção compartilhem das mesmas unidades durante a comunicação.

Nesta seção, nos detivemos apenas a apresentar o conceito de paridade. Na seção seguinte (3.6 do Capítulo 2), traremos à luz como o conceito foi incorporado na proposta de Goldstein & Fowler (2003).

3.6 Uma proposta unificada para a produção e a percepção da fala

Neste nosso percurso dentro de uma abordagem dinâmica sobre a fala, temos como saldo as seguintes propostas: há a adoção do gesto articulatório como uma unidade de análise da produção da fala com a proposta da FAR (Browman & Goldstein, 1992); adota-se também o gesto como primitivo da percepção a partir de uma perspectiva realista e direta (Fowler, 1986); e, por fim, Liberman & Whalen (2000) propõem as exigências de paridade que todo sistema comunicativo deve ter. Como, então, unificar uma teoria fonológica, de produção e de percepção da fala?

A proposta de unificação do tratamento sobre a fonologia das línguas e a produção e percepção da fala de Goldstein & Fowler (2003) só é possível devido à adoção de uma abordagem de uso público da linguagem. Trata-se, pois, de um tratamento realista sobre a linguagem, uma vez que se espera que as entidades utilizadas para criar significados na linguagem sejam “tipos de ação pública” sem serem “exclusivamente categorias mentais” tal como adotadas pela maioria das teorias de fonologia, produção e percepção (Goldstein & Fowler, 2003: p.159). Para eles,

“uma teoria fonológica, então, deveria ser uma teoria

sobre as propriedades daquelas ações públicas, uma teoria de produção da fala deveria ser sobre como aquelas ações são alcançadas, e uma teoria de percepção da fala deveria ser sobre como as ações são percebidas” (Goldstein & Fowler, 2003: p.159)¹⁵⁰.

Como podemos notar nessas primeiras palavras dos autores, temos os três pontos desenvolvidos dentro de um arcabouço dinâmico sobre a linguagem: a teoria fonológica que descreve as propriedades das ações dos articuladores é a FAR (Browman & Goldstein, 1992); uma teoria de produção de fala que mostra como os primitivos são planejados e implementados no ato da fala é a Dinâmica de Tarefa (Saltzman & Kelso, 1993); e a teoria que propõe como as ações são percebidas é a TRD (Fowler, 1986). Ainda mais, segundo os autores, esses três pontos básicos que unificam o tratamento sobre a linguagem só são possíveis, porque há a adoção de uma mesma unidade em comum para realizar as combinações fonológicas necessárias em um dado sistema, bem como a sua implementação e a sua percepção., i.e.,

“a ideia central que eles compartilham é que deve haver uma moeda fonológica comum entre os conhecedores das formas da língua, os produtores e os percebedores delas. Ou seja, as formas da língua que os usuários conhecem, produzem e percebem devem ser as mesmas” (Goldstein & Fowler, 2003: p.172-173)¹⁵¹.

Essa moeda de circulação comum entre a fonologia, a ação e a percepção é o gesto articulatório. Primeiro, trata-se de uma unidade fonológica, pois os gestos podem ser pensados como um sistema combinatório que promove distinções dentro do léxico das línguas humanas. É também válido pensar que essa propriedade combinatória não é exclusiva dos fatos relativos à linguagem, uma vez que “outros sistemas naturais, tais como os compostos químicos e a recombinação genética”, (Goldstein & Fowler, 2003: p.160) compartilham dela. Ainda mais, com relação à natureza fonológica do gesto, é preciso reconhecer que “as unidades atômicas são distintas discretamente umas das outras, e elas mantêm suas descrições quando

¹⁵⁰ “A theory of phonology, then, should be a theory about the properties of those public actions, a theory of speech production should be about how those actions are achieved, and a theory of speech perception should be about how the actions are perceived” (Goldstein & Fowler, 2003: p. 159).

¹⁵¹ “[...] a core idea that they share is that there must be a common phonological currency among knowers of language forms, producers of the forms and perceivers of them. That is, the language forms that language users know, produce and perceive must be the same” (Goldstein & Fowler, 2003: p. 172-173).

combinadas para formarem novos objetos” (Goldstein & Fowler, 2003: p.160).

Segundo, trata-se de uma unidade de ação, pois, diferentemente das abordagens anteriores (herdeiras de um arcabouço gerativista) que propunham que a informação fonológica não era parte da produção, já que se tratava de um alvo “escondido na mente do produtor”, “as unidades gestuais são modeladas como sistemas dinâmicos” (Goldstein & Fowler, 2003: p.166) a partir de equações que expressam como o estado do sistema muda no tempo. O ganho com essa abordagem é a possibilidade de acompanhar os gestos diretamente no momento em que estão em movimento, uma vez que a variável tempo é integrada na sua especificação.

Terceiro, trata-se de uma unidade de percepção, pois, de acordo com a proposta de Gibson (1966),

“mídia informacional tal como a luz para a visão e a energia acústica para a audição de fato especificam as suas fontes distais. Isso ocorre pelo fato de haver uma relação causal entre as propriedades distais dos objetos e os eventos e estruturas nessas mídias, e devido às diferentes propriedades dos objetos e eventos tenderem a estruturar a mídia distintivamente. Desta maneira, a estrutura na mídia transmitida aos sistemas sensoriais serve como informação para a sua fonte distal. Nesta proposta, os percebedores detectam distribuições de energia padronizadas que estimulam seus órgãos do sentido (o 'estímulo proximal'), e percebem as propriedades e os eventos do mundo (eventos 'distais') que promovem as informações sobre as distribuições de energia” (Goldstein & Fowler, 2003: p.182)¹⁵².

Transpondo essa proposta para a percepção da fala, temos (cf. Fowler, 1986) que os gestos são os eventos distais que promovem a estruturação do sinal acústico da fala (o estímulo proximal). Essa proposta permite um olhar unificado sobre as várias modalidades de percepção, uma vez que os mesmos gestos podem estruturar a luz, por exemplo, e “isso dá origem à moeda comum de troca que permite que crianças e adultos integrem informação sobre a fala em várias modalidades”

¹⁵² “[...], informational media such as light for seeing, and acoustic energy for hearing do specify their distal sources. This is because there is a causal relation between properties of distal objects and events and structure in those media, and because distinctive properties of objects and events tend to structure the media distinctively. In this way, structure in media imparted to sensory systems serves as information for its distal source. In the account, when perceivers detect patterned energy distributions that stimulate their sense organs (the 'proximal stimulus'), they perceive the properties and events in the world ('distal' events) that the energy distributions provide information about” (Goldstein & Fowler, 2003: p.182).

(Goldstein & Fowler, 2003: p.183).

Além dos argumentos já apresentados em Fowler (1986 – tais como os achados de Liberman e colegas (1957, 1967) que mostram que a percepção segue a articulação quando o sinal acústico e a articulação vão em direções opostas, e os achados de McGurk & McDonald (1976) e Meltzoff & Moore (1982 *apud* Goldstein & Fowler, 2003: p.187) sobre a percepção audiovisual – Goldstein & Fowler (2003) mostram que há um mecanismo que embasa a proposta de moeda comum entre a produção e percepção da fala. Trata-se da descoberta sobre os neurônios-espelho na área F5¹⁵³ do córtex de macacos (cf. Rizzolatti, Fadiga, Gallese & Fogassi, 1996 *apud* Goldstein & Fowler, 2003: p.188). Foi verificado que tais neurônios são ativados quando os macacos veem e imitam uma ação realizada por outro macaco ou humanos. Esse mecanismo, então, pode ser o que conecta a ação e a percepção se levarmos em conta a proposta apresentada pelos autores. Eles ainda argumentam que “de fato, é notável que os neurônios-espelho sejam encontrados no macaco na área homóloga à área de Broca do cérebro humano, uma área que é ativada quando a linguagem é usada” (Goldstein & Fowler, 2003: p.188).

Embora esse argumento seja o mais atual e “surpreendente”, há uma questão que o achado do neurônio-espelho não consegue responder: como um neurônio consegue reconhecer uma ação como sendo a mesma? Para Goldstein & Fowler, os neurônios têm que possuir o conceito de moeda comum tal como proposto para o sistema humano de produção e percepção da fala. “Em suma, os neurônios-espelho devem constituir o cume de operações de um mecanismo neural inteligente, trabalhos dos quais não entendemos tanto quanto entendemos das realizações dos percebedores/atores” (Goldstein & Fowler, 2003: p.188¹⁵⁴)

Como podemos notar, ao se assumir uma perspectiva de uso público da linguagem, fica clara a proposta de que o gesto articulatório é uma unidade tanto abstrata quanto de ação e de percepção. É através da sua adoção como primitivo fonológico, acional e perceptual e da adoção das exigências de paridade que temos

¹⁵³ A área F5 (também chamada de área pré-motora ventral) está localizada no lobo frontal e foi nessa região que Rizzolatti e colegas (1996) encontraram esses neurônios que disparam quando o animal realizava uma tarefa motora específica (como pegar algo com a mão) ou quando ele via alguém realizar a mesma tarefa (cf. Lameira, Gawryszewski & Pereira Jr, 2006).

¹⁵⁴ “In short, mirror neurons must constitute the culmination of the operations of a very smart neural mechanism the workings of which we understand no better than we understand the achievements of perceivers/actors” (Goldstein & Fowler, 2003: p. 188).

uma proposta unificada sobre a comunicação humana. Ainda mais, é com a adoção do gesto como primitivo que é possível realizar um tratamento multimodal para a percepção, uma vez que se percebe o gesto, não há um único meio pelo qual os sentidos são estimulados. Percebe-se o gesto através do ar, luz ou tato. Essa abordagem dá conta, por exemplo, dos dados do *Efeito McGurk*, pois trabalha justamente com a integração dos sentidos, da sua multimodalidade.

3.7 Sumário da seção

Nesta seção, apresentamos o percurso da TRD. Esperamos que tenha ficado claro que a teoria foi proposta em decorrência dos dados que pareciam interessantes de serem tratados, i.e., foi a partir dos primeiros achados sobre aspectos coarticulatórios da percepção da fala que a autora se viu inclinada a adotar uma abordagem realista sobre a percepção (Gibson, 1966). Esse é um ponto relevante para a diferenciação da TRD com a TM, pois, enquanto a TM parte de aspectos de natureza simbólica e estática em seus experimentos, a TRD tenta verificar como a coarticulação colabora com a percepção. Isto é, a autora assume que percebe-se fala coarticulada (a produção real) e não símbolos que podem ser ou não ser implementados.

É importante também ressaltar que é com o desenvolvimento da FAR e com o conceito de paridade (Liberman & Whalen, 2000) que a TRD ganha cobertura empírica maior, bem como previsão explicativa. Tais aspectos desembocam na proposta de Goldstein & Fowler (2003) que unifica o tratamento da produção e percepção da fala em um arcabouço dinâmico.

Em suma, mostramos a partir do percurso histórico da TRD que, em uma abordagem dinâmica sobre a fala, é possível unificar o tratamento da produção e da percepção fazendo uso do gesto articulatório como unidade comum entre os dois processos.

CAPÍTULO 3: UMA POLÊMICA SOBRE A NATUREZA DOS PRIMITIVOS DE PERCEPÇÃO DA FALA

"– Um POP é uma coisa que não podemos ver, ou não vemos, ou nosso cérebro não nos deixa ver porque pensamos que é um problema de outra pessoa. É isso que POP quer dizer: Problema de Outra Pessoa. O cérebro simplesmente o apaga, como um ponto cego. Se você olhar diretamente para ele, não verá nada, a menos que saiba exatamente o que é. A única chance é conseguir ver algo olhando de soslaio."

Ford Prefect (Adams, D, 2010: p.23)

Nos capítulos anteriores, dedicamo-nos a apresentar (no Capítulo 1) as diferentes maneiras de se conceber a natureza dos primitivos de percepção da fala. De modo geral, os trabalhos que tentaram responder a essa questão a consideraram ou como um primitivo de análise composto por propriedades estritamente acústicas ou como um de base articulatória.

Por outro lado (no Capítulo 2), tentamos mostrar que a adoção de cada um desses primitivos perceptuais está vinculada à proposição de novas teorias fonológicas que, a seu tempo, sugerem novos primitivos para o tratamento do nível fonológico das línguas. Vale ainda salientar que tais teorias fonológicas estão relacionadas a Programas de Investigação Científica distintos (cf. Lakatos, 1979) que promovem olhares diferentes sobre quais são os objetos de interesse para cada uma das teorias que tratam dos aspectos fônicos da linguagem.

Por exemplo, no período Estruturalista, há uma busca pela invariância supostamente presente no sinal acústico, uma vez que o sinal de fala teria a função de transmitir informações de um emissor a um destinatário que atuam em um dado sistema de comunicação. É desse período a proposição de que os fonemas eram compostos por feixes de traços de base acústica (Jakobson, Fant & Halle, 1952) que promoviam distinção fonêmica no sistema fonológico das línguas. Com isso, esperava-se que tais propriedades acústicas estivessem presentes na realização dos fones e que elas fizessem com que os aspectos distintivos de um dado fonema fossem perceptualmente audíveis.

Diante, então, do fracasso da busca pela invariância acústica (tal como apresentado por Liberman – 1957), os pesquisadores que se debruçavam sobre os estudos de percepção da fala começaram a estipular que a sua natureza era de base articulatória e não acústica. Ou seja, se o domínio acústico não apresenta propriedades invariantes para compor os fonemas, o domínio articulatório deveria ser invariante. Caso contrário, seria impossível a transmissão de informação em um dado sistema comunicativo.

Desde a proposição da *Early Motor Theory* (Liberman, 1957), os trabalhos de Liberman e colaboradores nos Laboratórios *Haskins* precisavam de um arcabouço teórico que possibilitasse o desenvolvimento da proposta de que a percepção era guiada pela articulação. Tais marcos teóricos (a Fonologia Gerativa Padrão de Chomsky & Halle -1968 – e a Modularidade da Mente de Fodor – 1983) só surgiram anos após esses primeiros achados e isso possibilitou a formulação da TM (Liberman & Mattingly, 1985). Nessa teoria, os primitivos perceptuais são de base articulatória abstrata, i.e., trata-se de primitivos que representam os comandos motores pretendidos pelos falantes que são recuperados através de um módulo fonético que é alimentado pelos sons produzidos e veiculados no ar durante a fonação.

A terceira teoria de percepção da fala (a TRD – Fowler, 1984 e seguintes) começa a ser ensaiada a partir dos avanços promovidos nas técnicas de análise acústica e articulatória da fala. Tais avanços possibilitam que se tente buscar como as características coarticulatórias presentes na produção dos sons influencia a percepção de vogais e consoantes. Nessa perspectiva, assume-se a coarticulação como uma característica inerente à fala e que ela estaria envolvida na sua percepção. Trata-se de uma abordagem bem distinta das anteriores, pois toma a fala dinamicamente, i.e., a fala passa a ser considerada no seu desenvolvimento no tempo.

A incorporação do tempo nas análises sobre a fala possibilitam a adoção de um novo primitivo de análise fonológica (o *gesto articulatório* – Browman & Goldstein, 1992), que, por sua vez, fomenta que se olhe para a percepção de uma maneira não puramente simbólica (tal como na TM), mas sim realista (com base em movimentos que de fato acontecem no momento da produção da fala) e dentro de

um sistema de desenvolvimento ecológico da espécie humana (Gibson, 1966).

Como se pode notar, em cada um desses Programas de Investigação Científica (a saber, Estruturalista, Gerativista e Dinamicista), há a proposição de teorias de Fonologia e de produção e percepção da fala que sustentam seus respectivos núcleos epistemológicos.

A esta altura deste trabalho, especulamos que tais questões acerca das diferenças entre as teorias de percepção da fala (e dos seus vínculos aos programas de investigação científica) não estejam totalmente claros. Saliento um motivo para isso.

A maneira como foram apresentadas as teorias no Capítulo 2 pode ter dado a falsa impressão de que a proposição de cada uma das teorias de percepção da fala tenha ocorrido de forma linear. Em outras palavras, poderíamos pensar que uma teoria estaria superando e substituindo outra (tal como queria Kuhn - 1975). Não se trata disso. Apresentamos a constituição de cada teoria isoladamente e vinculada a um programa de investigação científica em linguística para que não houvesse dúvidas com relação aos domínios epistemológicos de cada proposta.

No entanto, as teorias de percepção da fala se encontram em concorrência, uma vez que não há unanimidade com relação a qual seria a natureza dos primitivos perceptuais. Isso se verifica com a polêmica na qual Fowler (1996) e Ohala (1996) se envolvem. Esse é motivo pelo qual apresentaremos essa querela neste capítulo. Nas subseções que se seguem, iremos indicar os aspectos teóricos que impossibilitam o fechamento da polêmica de modo a sustentar a incompatibilidade epistemológica entre essas teorias.

Primeiro, passaremos à apresentação da proposta e dos argumentos de cada um dos envolvidos na polêmica. Segundo, realizaremos um exercício de responder ao debate a partir das premissas da TM para ilustrar que cada uma das teorias considera a percepção da fala de uma maneira distinta. Terceiro, buscaremos mostrar que cada uma dessas teorias considera os fatos acústicos e articulatórios de uma maneira diferente. Por fim, sumaremos os motivos pelos quais a polêmica não fecha.

1 UMA POLÊMICA SOBRE A NATUREZA DOS PRIMITIVOS DE PERCEPÇÃO DA FALA

Nesta seção apresentaremos uma polêmica em torno da natureza da percepção da fala. Trata-se do posicionamento de base auditiva (ou de base acústica) defendido por Ohala (1996) contra as propostas de base articulatória de Fowler (1996). Após apresentar a polêmica, realizaremos uma análise das propostas salientando a diferença que há entre cada uma das teorias de percepção da fala bem como mostraremos o papel que o acústico assume em cada uma delas. O ponto de chegada da nossa proposta é mostrar que a polêmica não se desenvolve justamente pelo fato de os programas de pesquisa defenderem seus pressupostos teóricos a partir de um local específico bem demarcado.

A escolha dessa polêmica em nossa análise não foi tomada ao acaso. Primeiramente, é preciso ressaltar que as discussões acerca da natureza da percepção da fala têm seu início na década de 1950 com os primeiros achados de Liberman e colegas nos Laboratórios *Haskins*. No entanto, as discussões não eram “travadas” em forma de “troca de textos”. Eram antes de tudo discussões acerca de achados experimentais que colocavam em xeque hipóteses (ou concepções) ainda insipientes sobre a percepção da fala. Basta lembrarmos dos relatos e das discussões sobre o fato de que os “invariantes” acústicos da fala eram determinados na base da “tentativa-e-erro” a partir do auxílio do *Pattern Playback*. Até a proposição da TM, as discussões sobre a natureza da percepção da fala eram realizadas sem ter um “oponente” bem determinado. Só por esse aspecto, a troca de textos entre Ohala e Fowler já mereceria atenção.

Outro motivo pelo qual tomamos esse “debate” como objeto de estudo é o fato de se tratar de um discurso polêmico sobre a natureza da percepção da fala, i.e., trata-se do ambiente ideal para se opor ideias e se disputar “território” para uma determinada teoria. Entretanto, o ponto interessante dessa polêmica é justamente o fato de ela ser colocada e não se desenvolver. A nossa hipótese é a de que a discussão não ganha atenção pelo fato de as teorias estarem assentadas sobre terrenos epistemológicos distintos, fazendo com que a polêmica não chegue a lugar

algum¹⁵⁵.

Passemos, então, para a polêmica e depois para a sua análise e discussão sobre o motivo pelo qual a polêmica na fecha.

1.1 Entre o acústico e o articulatório: Ohala *versus* Fowler

Ohala (1996) argumenta em seu texto que “a percepção da fala não envolve a recuperação das articulações que geram o sinal acústico” (Ohala, 1996: p.1718). Para o autor, recuperar a articulação subjacente à fala não é uma tarefa essencial para entender a sua percepção. Em princípio, ela é importante apenas para o trabalho do *cientista da fala*. Tendo isso em vista, ele vai sugerir que as teorias de base motora – tais como a Teoria Motora de Percepção da Fala de Liberman & Mattingly (1985) e a Teoria do Realismo Direto da Percepção da Fala de Fowler (1986) – são uma maneira extravagante de encarar a percepção da fala.

Seus argumentos em favor de um olhar auditivo sobre a percepção levam em conta: 1) algumas evidências fonológicas; 2) mudanças sonoras; e 3) algumas reflexões que indicam a falta de plausibilidade dessas teorias. Vamos a elas.

Fonologicamente há características nos sons das línguas que são melhor capturadas por traços de base acústica¹⁵⁶, tais como [grave] e [flat] na proposta dos *Preliminaries of Speech Analysis* (Jakobson, Fant e Halle, 1952). Sons [+graves] apresentam predominantemente energia de produção localizada na porção mais baixa e inicial do espectro. Por exemplo, sons labiais e velares seriam identificados como [+grave]; em oposição, apicais e palatais não. Tais dados mostram que há “um

¹⁵⁵ Há pelos duas outras maneiras de se considerar este quadro apresentado. Uma é considerar que a discussão sobre o tema não se desenvolve do mesmo modo que a polêmica. De certa forma, isso explicaria a nossa sensação de que a polêmica não se desenvolve mas as teorias, sim. Um outro aspecto seria o fato de as discussões sobre as fronteiras entre as disciplinas Fonética e Fonologia serem ainda muito recentes. Podemos dizer que, salvo alguns espaços simpatizantes (cf. Albano, 2001, Silva, 2002), a Fonética ainda permanece na periferia do núcleo duro das questões linguísticas.

¹⁵⁶ Poderíamos nos perguntar se os traços [grave] ([bemo]) e [flat] são de base acústica ou auditiva. Esse ponto não fica muito claro para a proposta dos *Preliminaries*, uma vez que o tratamento instrumental para a determinação dos traços a partir das características visuais (via espectrografia) do sinal acústico é realizado para validar o que se ouve como distintivo. Talvez seria melhor considerá-los como acústico-auditivo, já que a percepção é a audição das características acústicas que promovem distinções fonológicas.

conjunto descontínuo¹⁵⁷ de um ponto de vista articulatorio” (Ohala, 1996: p.1721). O mesmo traço consegue explicar por que motivo algumas vogais do inglês não sofreram alteração no *English Vowel Shift*. Por exemplo, historicamente o ditongo [aw] do inglês veio de um [u:] (e.g., *mouse* [maws] < [mu:s] e *south* [sawθ] < [su:θ]), mas hoje o ditongo [aw] apenas ocorre antes de sons que têm o traço [-grave] (e.g., *shout*, *loud*, *house*, *howl*, *town*) e nunca antes de [+grave] (e.g., *[awp], *[awm], *[awk]). A razão pela qual isso ocorre é devido à possibilidade de consoantes [-grave] produzirem um traçado descendente do F2 com relação à vogal [u]. Em contrapartida, consoantes [+grave] não produzem essa “queda” de F2 pelo fato de possuírem um F2 similar ao de [u], não conseguindo criar a vogal ditongada. Ou seja, Ohala argumenta em favor de uma bordagem acústica pois a percepção de um ditongo está associada ao traçado de F2 e não com as articulações envolvidas na sua produção.

Na mesma esteira de descontinuidade do ponto de vista articulatorio está o traço [flat]. Ele se aplica aos sons retroflexos e aos que apresentam articulação secundária nos lábios, véu (e úvula) ou faringe. Trata-se também de um conjunto de articulações descontínuas, pois não se incluem apicalizações ou palatalizações.

Ainda dentro da gama de evidências fonológicas apontadas por Ohala, há o argumento de que, como os sons são veiculados na fala, espera-se que as diferenciações sejam encontradas nesse domínio. Para endossar essa hipótese, o autor aponta para os achados de Stevens (1989) e Ohala (1983), que diz:

“Parece que os inventários dos sons da fala exploram uma relação não linear entre a produção da articulação e dos sons para fazer com que os elementos dos sons da fala se diferenciem” (Ohala, 1996: p. 1719)¹⁵⁸.

Os argumentos sobre a mudança sonora nas línguas do mundo caminham para a mesma direção de dizer que elas ocorrem por haver alguma variação no

¹⁵⁷ Ohala considera o traço [+grave] como articulatoriamente descontínuo porque ele especifica regiões articulatorias distintas (e opostas) do trato vocal. Cabe-nos perguntar qual foi o motivo que levou Ohala a iniciar a sua análise pelos traços positivos, uma vez que eles poderiam categorizar a porção medial do trato, que caracterizaria um contínuo articulatorio. A mesma reflexão cabe para o traço [flat]. Ohala não discute tais motivos e tampouco desenvolve as questões subjacentes referentes à classe de som. Contudo, como consideramos que se trata de um problema para todas as teorias fonológicas, decidimos não entrar nessa discussão aqui.

¹⁵⁸ “The answer is that speech sound inventories exploit the nonlinear relation between articulation and the sound produced in order to make the elements of speech *sound* different (cf. Ohala, 1983; Stevens, 1989)” (Ohala, 1996: p.1719).

aspecto acústico-auditivo e não na articulação. Em seu texto há a apresentação de dados sobre: 1) as labiais palatalizadas (do tcheco padrão, italiano de Roma e proto-bantu) que são confundidas com alveolares (realizadas, respectivamente, como forma inovadora no boêmio do leste, italiano genovês e zulu) – cf. Ohala 1978 *apud* Ohala, 1996: p. 1722; 2) a nasalização espontânea do sânscrito para o hindi e do francês para o bretão (Ohala & Ohala, 1993 *apud* Ohala, 1996: p.1722); e 3) as nasais epentéticas antes de oclusivas sonoras na mudança do hindi antigo para o moderno.

Por fim, há os argumentos contra a falta de plausibilidade das teorias que propõem que a percepção da fala é guiada pela articulação. Ohala argumenta que há estudos com animais e com crianças em período anterior à fala que mostram a sua capacidade de categorizar sons da fala. Para ele, é muito difícil que nesses casos esteja ocorrendo a recuperação das articulações utilizadas.

Ainda mais, Ohala se pergunta se conseguimos recuperar todas as articulações utilizadas quando ouvimos uma corneta tocando, uma vez que seu som é produzido a partir da variação de pressão da respiração e da tensão dos lábios¹⁵⁹.

Um outro ponto interessante apontado pelo autor é o fato de aprendizes de segunda língua e pessoas em terapia de fala aprenderem a produzir os sons de uma língua a partir de explicações articulatórias. No entanto, tais tarefas tomam muito tempo e esforço, sugerindo que a aprendizagem de sons talvez não se dê através do conhecimento articulatório e sim acústico-auditivo.

Finalmente, há um último comentário do autor denominado *Chicken Little Inquiry*:

“Chicken Little (ou em algumas versões, Henny Penny) é o personagem principal de um desenho infantil familiar à maioria dos falantes de inglês. Um dia, Chicken Little recebeu um grande sopro na cabeça e então colocou todo o celeiro em pânico com a sua teoria de que o céu estava caindo. A resolução da estória não envolve experimentos que testam se o céu, de fato, estava caindo; envolveu uma investigação sobre a razão pela qual Chicken Little

¹⁵⁹ Talvez consigamos sim recuperar essa variação de pressão da respiração e pressão dos lábios quando ouvimos uma pessoa tocando uma corneta. Basta pensarmos que, de modo geral, conseguimos imitar os sons de um instrumento dessa natureza. A destreza na afinação e fidelidade de timbre depende do imitador da mesma forma que o som do instrumento depende do músico que está tocando.

pensava que o céu estivesse caindo. A evidência imediata foi um inchaço no topo de sua cabeça. Onde estava ele quando o dano ocorreu? Embaixo de um carvalho. E eis que uma grande bolota foi encontrada no mesmo local! A moral da história - que não ficou tão explícita assim - é de que, antes de investir muito tempo e esforços em experimentos caros para avaliar uma hipótese qualquer, devemos primeiro focalizar a motivação para a hipótese construída. Então devemos nos perguntar se as observações que originaram a hipótese poderiam ser tratadas por outras hipóteses menos extravagantes” (Ohala, 1996: p.1724)¹⁶⁰.

Com essa anedota, o autor conclui que as teorias de base motora são em realidade uma hipótese extravagante sobre o fato de haver falta de invariância entre as unidades linguísticas assumidas na fala e sua manifestação acústica. Para Ohala, esse argumento é do tipo “Como não encontro isso aqui, ele deve estar ali!”.

Ainda mais, ele sugere que “as invariantes estão presentes no sinal acústico, as pessoas apenas estão olhando para os traços errados” (cf, Stevens & Blumstein 1978 *apud* Ohala, 1996: p.1724) e que “algumas relações de ordem alta entre certos parâmetros acústicos mostrarão uma relação invariável com as unidades linguísticas” (Sussman *et al* 1991 *apud* Ohala, 1996: p.1724). Para ele, não há um único traço invariante para cada unidade. Ao contrário, há uma coleção de traços e a identidade de uma dada unidade é determinada se algum número crítico desses traços estiver presente.

Na tentativa ainda de engrossar os argumentos, ele sugere que as teorias motoras estão considerando os fonemas como unidades de codificação; de modo que difones poderiam ser candidatos melhores.

A resposta de Fowler (1996) a Ohala tem três objetivos: 1) mostrar que há diferenças entre a Teoria Motora (TM) e a Teoria do Realismo Direto (TRD), pois ele as considerou como equivalentes; 2) apresentar argumentos a favor da percepção

¹⁶⁰ “Chicken Little is the principal figure in a children’s story familiar to most English speakers. Chicken Little received a tremendous blow on the head one day and then set the entire barnyard into a panic with her theory that the sky was falling. The resolution of the story did not involve experiments testing whether the sky was, in fact, falling; it involved an inquiry as to why Chicken Little *thought* it was falling. The immediate evidence was a swelling on the top of her head. Where was she when the injury occurred? Under an oak tree. A large acorn was found on the very spot! The moral of the story—not stated quite so explicitly—is that before investing a lot of time and effort in costly experiments evaluating a given hypothesis, we should first look at the motivation for the hypothesis having been made. Then we should ask whether the observations prompting the hypothesis might be accounted for by other, less extravagant, hypotheses”. (Ohala, 1996: p. 1724).

dos gestos articulatórios; e, por fim, 3) trazer argumentos que digam que os pontos levantados por Ohala não refutam a teoria. Com isso, passaremos à apresentação dos pontos principais de cada objetivo.

Basicamente, o desenvolvimento da TRD está ligado com o desenvolvimento das teorias de produção da fala e fonologia (como a Fonologia Articulatória de Browman & Goldstein, 1992) e com uma teoria mais geral de percepção direta (Gibson 1966, 1979). Para esses estudiosos deste viés teórico, a percepção dos gestos não é logicamente necessária para ouvintes/falantes habilidosos, ela é biologicamente necessária. Isto é, a percepção da fala somente é possível se ela for direta (sem intermediários transformacionais) e se considerar que há um sistema perceptual com funções universais moldados pela seleção natural.

Esta última é biologicamente fundamental porque é por esse sistema perceptual que os animais conseguem reconhecer seu meio e isso é claro que foi moldado pela evolução. Ainda mais, os sistemas utilizam estruturas em mídia que é causada por eventos no ambiente como informação para esses eventos, i.e., luz para a visão, pele para o tato, e ar para a audição. Os seres, então, perceberiam os componentes de seu meio que causaram a estrutura.

Fowler mostra alguns achados experimentais como evidências para a sua proposta. Um deles é o chamado *Efeito McGurk* e vamos nos deter, por ora, apenas a ele. McDonald & McGurk (1976) sincronizaram a imagem da boca de um falante produzindo a sílaba /da/ com o sinal acústico de um /ma/. A maioria das pessoas (92% dos informantes) respondeu que a imagem havia produzido um /na/, i.e., os informantes perceberam o ponto de articulação visualmente e o modo de articulação auditivamente. Os autores sugeriram que as pessoas percebem gestos e alguns gestos são percebidos visualmente e outros auditivamente. Outro trabalho (Massaro 1987) sugeriu que em realidade os falantes associam as experiências (armazenadas na memória) de ver e ouvir as pessoas falando com uma imagem mental de um fonema (ou sílaba). Isso, para ele, explicaria o fato de as pessoas conseguirem ler lábios, por exemplo. No entanto, Fowler e Dekle (1991) conseguiram mostrar que tais associações não ocorrem.

Para tanto, eles partem do seguinte princípio: se o *Efeito McGurk* emerge da

associação na memória entre um sinal (ou gesto) e um som produzido por um falante, então um efeito análogo pode ocorrer quando ocorre a integração entre o tato (um sistema de base motora e articulatória) e a fala. Ou seja, os experimentadores queriam testar se o *Efeito McGurk* ocorre quando um sujeito (de olhos fechados) ouve fala e toca a face de alguém. Para testar essa hipótese eles realizaram dois experimentos.

No primeiro, eles sintetizaram um contínuo de sílabas de /ba/ até /ga/. A tarefa consistia em identificar a sílaba que eles ouviam e então identificar de alguma maneira a sílaba que eles viam, pois à medida que era apresentado um estímulo auditivo de /ba/ ou /ga/ era apresentada na tela do computador ou a sílaba BA ou GA. Nesse experimento, foi possível notar que há muito pouca relação entre o som e a sílaba impressa na tela do computador, mostrando que há pouca associação entre o estímulo auditivo e a memória visual de sua escrita.

No segundo, os experimentadores substituíram as sílabas impressas de BA e GA por estímulos manuais (táteis) impressos que simulavam faces produzindo a sílaba /ba/ ou /ga/. Ao contrário do primeiro experimento, neste segundo houve grande efeito das sílabas sentidas sobre as ouvidas. Para eles, então, esses resultados mostram que a abordagem por associação é incompatível com o *Efeito McGurk*, mas é tranquilamente acomodada dentro da TRD.

Fowler, em resposta a Ohala, mostra que as evidências fonológicas apresentadas por ele não falseiam a TRD. Quando Ohala diz que se espera que as diferenciações ocorram no sinal acústico, pois a fala assim é veiculada, não há o falseamento do argumento da DR de que em realidade os gestos possuem propriedades invariáveis e elas causam distinções no meio, no sinal acústico. Ela cita o fato de que quando há *gaps* nos inventários das línguas, quase sempre se trata de um /p/. Isso justamente ocorre porque /p/ possui *bursts* fracos, i.e., ele não consegue ser “bem” veiculado no ar.

Com referência às dúvidas anedóticas de Ohala, a autora apresenta alguns argumentos interessantes. Para as crianças, o resgate dos gestos do sinal de fala não é apenas biologicamente necessário, é logicamente necessário. Elas precisam recuperar os gestos para aprender a produzir a fala. Esse ponto, embora não esteja

claramente na resposta de Fowler, é endossado pelos achados do *Efeito McGurk* (seção 2.3.2, capítulo 2) que apontam para a preferência na recuperação de aspectos acústicos por crianças, ou seja, aprender a recuperar as articulações seria logicamente necessário para aprender a falar.

Na mesma esteira está o fato de que por seleção natural a percepção de eventos sonoros especificados pelo sinal acústico promove a sobrevivência da espécie com mais facilidade do que o faria ao ouvir ar estruturado por si só. Para ela, perceber o mundo baseado em uma estrutura em mídia é uma função evoluída, mas isso não pode depender de um mapeamento *built-in* de estruturas para eventos, i.e., a fala não seria especial como queriam os adeptos da TM.

Por fim, Fowler pergunta: cornetas e automóveis? Como isso refuta a teoria? Da mesma maneira ela se pergunta (ao considerar o critério *Chicken Little*): se a teoria é extravagante, qual seria a teoria menos extravagante? Para ela, não há!

Diante desse debate, notamos que a escolha de um primitivo para a percepção da fala não é ponto pacífico nos estudos fônicos. A nossa proposta é a de que a natureza dos primitivos de percepção da fala está vinculada à natureza dos primitivos fonológicos que, por sua vez, são sugeridos pelos cortes epistemológicos sugeridos pelas teorias fonológicas.

1.2 Fazendo a vez de advogado do diabo ou como a TM responderia à polêmica

Na seção anterior (1.1 do Capítulo 3), há a apresentação das hipóteses e dos argumentos sustentados por Ohala e Fowler com relação à polêmica sobre a natureza ou acústica ou articulatória da percepção da fala. Diante deles, nos perguntamos: quais seriam os motivos pelos quais a polêmica não se desenvolve?

Um dos motivos pode estar justamente vinculado à relação existente entre o fato de cada uma das teorias se fundamentar em seu próprio programa de investigação científica. Ao que parece, esse fechamento tem o objetivo de desenvolver pesquisas que promovam o avanço da teoria. Para tanto, esse olhar

estreito às premissas da teoria também promove o desinteresse¹⁶¹ (epistemologicamente orientado) pelas críticas e ataques dos proponentes de teorias concorrentes, i.e., esse “fechar de olhos” pode fazer com que os proponentes das teorias não se importem com as críticas e discussões levantadas pelos seus adversários, uma vez que tais comentários não seriam interessantes para o programa de investigação ao qual as teorias estão vinculadas.

Diante disso, o desinteresse pode ser definido como uma estratégia de não se importar com as críticas e com os desenvolvimentos teóricos e experimentais de seus adversários¹⁶². É possível verificá-lo na polêmica apresentada anteriormente na seção 1.1 (do Capítulo 3), pois Fowler não se “importa” com as críticas de Ohala. Por exemplo, Ohala apresenta evidências sobre a fonologia e as mudanças das línguas para sustentar a proposição de um primitivo de base acústica para a percepção, enquanto Fowler fundamenta sua teoria com os dados sobre o *Efeito McGurk*. Nenhum deles apresenta um mesmo fenômeno com outra interpretação. A evasiva predileta nesses casos parece ser a de que os dados apresentados não refutam a teoria, indicando uma falta de interesse sobre o que as críticas do adversário têm a dizer. Para Fowler, seu adversário sequer entendeu qual é o propósito da TRD, a ponto de ela crer necessário “esclarecer” pontos sobre como a TM e a TRD encaram a percepção dos sons da fala de maneira distinta. Sua estratégia é a de rebaixar seu crítico! Seria como dizer que o adversário *misturou alhos com bugalhos*.

Também é possível identificar um descompasso quanto aos objetivos de pesquisa formulados por cada um dos arcabouços teóricos. A busca pelos traços invariantes das articulações (tal como pretende a pesquisa dentro de um arcabouço dinâmico) não é de interesse de um programa calcado em propriedades acústicas.

Esperamos ter deixado claro (após o percurso traçado no Capítulo 2) que cada uma das teorias de percepção da fala caminha para uma direção que,

¹⁶¹ Não estamos tratando o *desinteresse* como sendo uma “teimosia” dos proponentes de teorias. Estamos considerando o desinteresse como uma estratégia epistemológica para que se consiga desenvolver uma teoria, i.e., interessar-se por toda crítica não promoveria o avanço da teoria, uma vez que seus proponentes a todo momento teriam que revisar suas premissas. Lakatos (1979) tratou dessa questão e a chamou de *princípio de tenacidade*.

¹⁶² Lakatos (1970) trataria dessa questão a partir do princípio de tenacidade. Trata-se da tendência dos programas de investigação científica continuarem a se desenvolver mesmo quando se deparam com contraexemplos “recalcitrantes”.

possivelmente, não se encontrarão em algum ponto, pois cada fato novo descoberto por cada um das teorias está em um lugar diferente do mundo. Trata-se de objetos teóricos e observacionais distintos. Seria como se cada uma das teorias tivesse uma lanterna diferente para iluminar as suas descobertas. Por exemplo, uma teoria usa uma que clareia apenas coisas verdes e outra, coisas laranja. Por si só, essas lanternas trariam à tona objetos distintos. No entanto, podemos notar que cada uma dessas lanternas está apontada para um ponto diferente da realidade. Uma delas é apontada para a água e a outra, para o céu¹⁶³. Obviamente, cada uma dessas teorias encontrará coisas distintas em lugares distintos.

É esta relação entre as bases epistemológicas nas quais cada uma das teorias são assentadas e a falta de interesse pela crítica dos adversários que caracteriza a dificuldade de se desenvolver a polêmica relacionada aos primitivos de análise.

Por exemplo, é por este motivo que, mesmo a fonologia estruturalista dos *Preliminaries* não tendo sido leva a cabo, e sendo substituída por abordagens gerativistas de fonologia e de percepção da fala (como na TM), teóricos como Ohala ainda buscam dados para verificar quais são os traços linguisticamente relevantes do sinal acústico que promovem distinções nas fonologias das línguas. Isto é, não se abandona a possibilidade de se encontrar argumentos relevantes para sustentar o fato de que os primitivos perceptuais são de base acústica.

Ao que parece, o abandono de uma teoria ocorre apenas quando há a diminuição das pesquisas dentro de seu programa de investigação que, por consequência, acaba perdendo terreno e seguidores. Este parece ser o caso da TM, porque nesta polêmica entre Ohala e Fowler faltou o posicionamento de Liberman, já que a própria Fowler coloca que muito pouco há de comum entre a TM e a TRD.

Para exemplificar que cada uma dessas teorias assenta suas propostas em terrenos epistemológicos distintos, decidimos fazer a vez de *advogado do diabo* e mostrar que a TM apresentava objetivos distintos da TRD e de uma abordagem que assume primitivos de base acústica.

Primeiramente, ao contrário das propostas de base auditivas de Ohala e de

¹⁶³ Esse ponto é uma adaptação da piada do bêbado apresentada em Borges Neto (2004).

base articulatória de Fowler, a TM propõe que a percepção da fala envolve a recuperação dos gestos pretendidos pelo falante. Segundo, a TM assume uma abordagem simbólica sobre a percepção da fala, ao contrário da proposta realista e direta da TRD. Não se trata das articulações que ocorrem, que são implementadas. Para a TM, as articulações recuperadas são os comandos pretendidos pelo falante.

Em outras palavras, a TM parte da assunção de que a percepção recupera os gestos pretendidos e por isso não encontramos uma relação de um para um entre as características acústicas e as articulações realizadas no momento de produção da fala.

Para sustentar sua proposta, a TM apresenta evidências experimentais de que não há relação de um para um entre o sinal acústico e a articulação dos sons da fala. As evidências apresentadas são os achados da década de 1950 com o auxílio do *Pattern Playback* (Liberman e colegas, 1952 e seguintes), nos quais não se encontravam características acústicas (transição de F2) para o ponto de articulação de consoantes. Ao que tudo indica, as transições assumem valores distintos a depender da vogal do núcleo silábico.

Outra evidência apresentada para proteger as propostas da TM é o fato de a percepção seguir a articulação quando há um desencontro entre a articulação e os sons. Essa assunção é ilustrada pelo *Efeito McGurk*. Para a TM, o *Efeito* é uma prova cabal sobre a existência de módulos que, dado o desencontro entre as articulações e os sons promovidas pelo ambiente experimental do *Efeito*, a percepção se guia pela articulação. Ou seja, mesmo depois de ter conhecimento sobre o *Efeito* e de como o experimento foi conduzido, os sujeitos não conseguem “desfazer” a ilusão. Para a TM, trata-se de uma operação mandatória do módulo fonético que executa ações que recupera os estímulos distais sem se importar com o que acreditamos ou sabemos.

No que concerne ao que seria interessante de se pesquisar dentro da TM, podemos notar que os estudos desenvolvidos estavam organizados com relação ao seu vínculo com a *modularidade da mente* (Fodor, 1983), pois se sugerirá pesquisas que endossem a *especialidade* da fala em seus vários níveis. Isto é, para o desenvolvimento da proposta era preciso pesquisar os aspectos que faziam da

linguagem algo *especial*. Tanto que Mattingly & Liberman (1985) argumentam que a sua proposta sobre a percepção é totalmente compatível com os achados sobre a sintaxe, por exemplo. Com relação à percepção (o nosso foco de pesquisa), os autores propõem o desenvolvimento de mais pesquisas sobre a percepção duplex que poderia dizer mais sobre como o *módulo fonético* funcionaria.

Como podemos ver, a proposta da TM é diferente das de base auditiva e o de base articulatória da TRD. Tanto a proposição de primitivos de análise perceptual quanto as evidências experimentais que sustentam o modelo são distintos. Em outras palavras, enquanto a abordagem auditiva busca por invariantes no sinal acústico e a TRD tenta determinar quais são os invariantes articulatórios que são transmitidos por uma mídia informacional (o ar para a fonação e a luz refletida para as línguas de sinais, por exemplo), a TM busca determinar quais são os aspectos que fazem a fala um “instrumento” especial. Ou seja, para a TM os argumentos de Ohala não invalidariam a proposta motora uma vez que as evidências fonológicas e de mudança das línguas do mundo apresentadas poderiam mostrar que as alterações acústicas são detectadas pelo módulo fonético como estímulos proximais e que a percepção (e em seguida a mudança linguística) estaria rastreando os estímulos distais (a articulação). É preciso lembrar que tais articulações são pretendidas, de modo que os ouvintes perceberiam quais eram as intenções do falante, i.e., recuperar-se-ia a forma subjacente, abstrata.

Ainda mais, os argumentos de Ohala com relação aos sons de motores e a recuperação das articulações que promovem as alterações de pressão do aparelho fonador de um trompetista seriam facilmente respondidos dizendo que não se recuperam as articulações envolvidas nesses contextos pois não se trata da fala. Sendo a fala especial, o módulo fonético não se daria ao trabalho de “tentar” recuperar as articulações de eventos acústicos que não são reconhecidos como fala.

Ao imaginarmos qual seria o posicionamento da TM diante da polêmica acerca dos primitivos de percepção da fala, não há como não se perguntar por que a TM não entrou na discussão. Ao nosso ver, no momento da polêmica, a TM já poderia ou ter perdido seguidores ou estar ocupada com outra agenda de pesquisa. Tal agenda poderia até mesmo estar relacionada a alguma revisão de pressupostos teóricos devido ao vínculo a outro programa de pesquisa científica, como o de base

dinâmica ou ecológica.

De qualquer forma, é preciso considerar também que pode ter havido o abandono da teoria que decorre da desistência de uma das partes (que é difícil ocorrer por vontade própria) ou pela morte de um de seus proponentes. Este parece ser o caso da TM. O pronunciamento hipotético da TM feito aqui mostra que a teoria já não apresentava explicações “convincentes” ou seguidores. Este segundo aspecto aproxima as teorias às doutrinas. Esta questão merece atenção, mas não vamos tratar dela aqui¹⁶⁴.

1.3 O que é o acústico e o articulatório para cada uma das teorias de percepção da fala?

Para tentar deixar mais clara a nossa hipótese de trabalho (a saber, as teorias de percepção da fala assumem primitivos perceptuais a depender da natureza dos primitivos fonológicos adotados nos programas de pesquisa científica a que estão filiadas epistemologicamente), apresentaremos nesta seção como cada uma das teorias considera os aspectos acústicos e articulatórios no processo de percepção da fala. Para realizar esta apresentação é importante não perder de vista que o fenômeno de *perceber a fala* deve ser visto como o mesmo para todas as abordagens. O que muda para cada uma delas é o estatuto que se dá para cada um dos processos envolvidos durante o processo perceptual. Diante disso, podemos dizer que todas as teorias têm que levar em conta que perceber a fala envolve pelo menos três processos físico-fisiológicos.

Primeiramente, é indispensável que tenhamos um falante que é considerado a fonte de emissão da fala. Este sujeito, com o auxílio de seus órgãos e articuladores, faz com que seu aparelho fonador ressoe e produza sons que, em segundo lugar, serão transmitidos pelo ar até os ouvidos de seu interlocutor. Terceiro, também se faz indispensável levar em conta que para se perceber a fala precisamos de um sujeito que receba essas vibrações do ar e que as interprete à medida que as ouve.

Essas condições são as mesmas para todas as teorias. Em outras palavras,

¹⁶⁴ Um agradecimento ao professor Kleber Candioto (PUC-PR) por me abrir os olhos a esta questão.

estamos dizendo que o fenômeno de perceber a fala é o mesmo para todas elas, uma vez que é inegável que há o envolvimento da movimentação de articuladores (uma contraparte articulatória) que promovem vibrações das partículas de ar que, por sua vez, são transmitidas (uma contraparte acústica) até os ouvidos de um receptor presente num dado momento comunicativo¹⁶⁵.

O que queremos apresentar nas linhas a seguir é que esse mesmo fenômeno de perceber a fala tem seus elementos tomados de maneira distinta a depender do programa de pesquisa a que estão filiados. Focaremos principalmente os papéis assumidos pelas contrapartes acústica e articulatória envolvidas no processo de percepção e produção da fala.

Por exemplo, no período Estruturalista, o meio acústico era visto como um canal de transmissão de informações. Nesse domínio, era esperado que os invariantes perceptuais fossem encontrados, pois assim o sistema fonológico de uma dada língua conseguiria representar as distinções suficientes e necessárias para a transmissão satisfatória de uma mensagem para um dado destinatário. Tais considerações sobre o conteúdo informativo (e distintivo) das mensagens estão presentes nas páginas iniciais dos *Preliminaries*:

“Qualquer distinção mínima carregada pela mensagem coloca o ouvinte diante de uma situação de duas escolhas. Dentro de um determinado idioma cada uma das oposições tem uma propriedade específica que a diferencia de todas as outras. O ouvinte é obrigado a escolher ou entre duas qualidades polares de uma mesma categoria, tais como grave *versus* agudo, compacto *versus* difuso, ou entre a presença e ausência de uma certa qualidade, tais como vozeado *versus* não vozeado, nasalizado *versus* não nasalizado, incisivo *versus* raso (plano)”. (Jakobson, Fant & Halle, 1952: p. 3)¹⁶⁶

Ainda mais, a procura por unidades mínimas distintivas (tanto para a

¹⁶⁵ É claro que é possível perceber fala através de outros meios, como via telefone, por gravações ou pela interceptação de uma mensagem que não foi endereçada diretamente aos nossos ouvidos (i.e., quando se escuta a conversa de outra pessoa). No entanto, o que não devemos perder de vista é que em todos esses casos alguém estava produzindo fala para alguém ouvir.

¹⁶⁶ “Any minimal distinction carried by the message confronts the listener with a two-choice situation. Within a given language each of these opposition has a specific property which differentiates it from all the others. The listener is obliged to chose either between two polar qualities of the same category, such as grave vs. acute, compact vs. diffuse, or between the presence and absence of a certain quality, such as voiced vs. unvoiced, nasalised vs. non-nasalized, sharpened vs. Non-sharpened (plain).” (Jakobson, Fant & Halle, 1952: p. 3).

produção quanto para a percepção da fala) era justificada pelo fato de os usuários da língua terem acesso exclusivamente ao meio acústico. Diante disso, esse domínio passa a ser o foco dos estudos do período, uma vez que a contraparte acústica do processo de produção da fala em um dado sistema comunicativo passa a ser considerada fundamental (e a única saída) para tratar a percepção.

É importante notar que ao se olhar exclusivamente para a mensagem, deixa-se de lado o ouvinte envolvido no processo perceptual. Melhor dizendo: o ouvinte possui um papel passivo na percepção da fala. Isto é, considerando que todas as distinções já estão presentes nesse domínio, o destinatário apenas recebe passivamente a mensagem através do meio (canal) acústico.

Por sua vez, ao nos debruçarmos sobre o domínio articulatório, é possível notar que desde o Estruturalismo já há a utilização do termo gesto. Entretanto, nesse período ele era tratado de outra maneira. E esse é o ponto interessante para a discussão realizada aqui. Em suma, falava-se de gesto muito mais como sinônimo de movimento, que, devido às propostas estruturalistas, não apresentava relevância linguística, i.e., não produzia distinções fonológicas. Isso se deve ao fato de que o movimento dos articuladores estava vinculado muito mais à Fonética do que à Fonologia, i.e., o gesto estava vinculado ao que se produzia e não ao que se acreditava realizar (Troubetzkoy, 1970[1939]).

De modo distinto do período estruturalista, o Gerativismo imprime um outro olhar para o sinal acústico e para as articulações. Embora Fonética e Fonologia apresentassem uma tímida tentativa de tratar aspectos fonéticos como fonológicos, o sinal acústico continua tratado como epifenômeno dentro do processo de produção e percepção da fala.

Do ponto de vista da produção da fala, o sinal acústico é o resultado físico de toda a implementação motora dos símbolos que compõem a gramática fonológica das línguas, i.e., trata-se do resultado da realização motora das regras simbólicas envolvidas nas transformações mentais e abstratas que precedem a enunciação. Do ponto de vista da percepção, o sinal acústico será o estímulo proximal que alimentará a mesma gramática fonológica a partir da audição. Em seguida, fazendo uso de transformações, o módulo fonético traduzirá os estímulos proximais para que

se recuperem os estímulos distais, i.e., o módulo fonético fará com que essa parte física (as ondas sonoras) seja convertida em símbolos de base articulatória. Para a TM, tais articulações sequer precisam ser realizadas, uma vez que o módulo fonético consegue recuperar o comando neuromotor (o gesto) pretendido.

Como se pode notar, ao contrário do estruturalismo, o usuário da língua do gerativismo possui papel importante no que concerne à implementação das regras armazenadas (como representações mentais) que geram a cadeia da fala. E estando esse falante em “posse” de um sistema que converte símbolos de base articulatória (tal como a proposta de Chomsky & Halle – 1968) em ondas sonoras, nada mais esperado que o ouvinte (um interlocutor num dado momento comunicativo qualquer), também “equipado” pelo mesmo sistema que converte símbolos em ondas sonoras e ondas sonoras em símbolos, consiga recuperar as articulações pretendidas pelo emissor de uma mensagem.

Em outras palavras, cabe ao interlocutor (o ouvinte) perceber o dado articulatório, pois o sinal acústico é considerado consequência da articulação. Esse aspecto confere ao domínio acústico um lugar secundário no processo de percepção da fala. Em suma, no gerativismo, há um falante parcialmente ativo, porque ele recebe o sinal acústico e o módulo fonético faz todo o trabalho de recuperar os estímulos distais (as articulações pretendidas) através da recepção e tradução dos estímulos proximais (as ondas sonoras) a partir de regras fonológicas. A citação de SPE salienta a preocupação de que a fonologia dê conta tanto da produção quanto da percepção, evidenciando uma tentativa de integrar alguns aspectos fonéticos na fonologia:

“Dada a estrutura superficial de uma sentença, as regras fonológicas de uma língua interagem com certas restrições fonéticas universais para derivar todos os fatos determinados gramaticalmente sobre a produção e a percepção desta sentença.” (Chomsky & Halle, 1968: p.293)¹⁶⁷

Seguindo essa lógica, ao pensarmos no domínio articulatório da produção e percepção da fala, a TM gramaticaliza o gesto, porque ele passa a fazer parte da fonologia da língua. Nesse modelo, o gesto (ainda concebido como o movimento

¹⁶⁷ “Given the surface structure of a sentence, the phonological rules of the language interact with certain universal phonetic constraints to derive all grammatically determined facts about the production and perception of this sentence” (Chomsky & Halle, 1968: p.293).

pretendido dos articuladores) assume as características estáticas, mentais e simbólicas compatíveis com a abordagem gerativo-transformacional de SPE.

Diante dos comentários acerca das diferenças entre as abordagens estruturalistas e gerativo-transformacionais sobre o lugar do acústico e do articulatório para cada um dos Programas de Pesquisa Científica, há a necessidade de abrir um breve parêntese para esclarecer um ponto sobre o posicionamento de Ohala (1996) dentro da polêmica travada com Fowler (1996). A maneira como organizamos o nosso texto pode ter sugerido que Ohala é um estruturalista, uma vez que o autor propõe primitivos perceptuais de base acústica.

Não se trata disso. A proposição de primitivos perceptuais de base acústica poderia ser acomodada em um arcabouço gerativista. Isso, talvez, facilitaria a tradução dos fenômenos acústicos físicos em símbolos, que seriam também de base acústica. No entanto, estamos aqui fazendo suposições acerca do posicionamento de Ohala¹⁶⁸. Em seus trabalhos, há uma clara posição particular de que não há como tratar de fenômenos puramente fonéticos ou puramente fonológicos, uma vez que os dois níveis são em realidade integrados. Em suas palavras:

“Minha visão é que entre fonologia e fonética, fonologia é a disciplina superordenada, não porque ela realiza mais ou está melhor desenvolvida – o oposto pode ser verdade – mas simplesmente porque ela observa e busca por respostas de fenômenos que envolvem o comportamento da fala de uma maneira mais abrangente. As respostas a essas questões surgirão, acredito, da fonética, da psicologia e dos estudos da cultura e sociedade – incluindo ambas as maneiras que esses domínios determinam o comportamento da fala atualmente e no passado. Embora neste trabalho eu tenha enfatizado a necessidade de traçar a fonética para questões fonológicas, eu não quero negligenciar os recursos nesses outros domínios. [...] Sem dúvida aqueles que usam o termo 'interface entre fonologia e fonética' creem que eles estão dando passos positivos para demolir a barreira entre os dois campos. Mas eu mantenho que se o termo 'interface' preserva a visão desatualizada com a qual estamos lidando largamente em campos autônomos, isso impedirá nossas chances de melhorar o passado. Eu proponho uma frase e um empreendimento alternativos: a

¹⁶⁸ Tal como feito neste trabalho sobre a trajetória dos estudos de Alvin Liberman, seria interessante historicizar as propostas de Ohala para que se tenha em conta mais claramente como se constituiu seu posicionamento sobre fonética e fonologia serem consideradas duas disciplinas integradas.

integração da fonologia e da fonética”. (Ohala, 1990: p.198)¹⁶⁹

Mesmo diante de uma proposta claramente sem dicotomizações, o posicionamento de Ohala não se adequa à unificação “radical” de Browman & Goldstein (1992), pois, embora o gesto articulatório seja de natureza numérica e simbólica, esse primitivo é, contrariamente à proposta acústica-auditiva de Ohala, de base articulatória e realista. Da mesma forma, a sua abordagem não nega a relevância de dados de natureza simbólica, pois basta lembrarmos dos dados acerca das fonologias das línguas em sua argumentação no debate. Entretanto, sua proposta parece direcionar os estudos fônicos para um embasamento laboratorial (ou experimental) dos dados que, por sua vez, podem colaborar para a adoção de primitivos fonológicos e perceptuais integrados, i.e., trata-se de uma fonética que responde a questões fonológicas.

Após fechar esse pequeno parêntese acerca do posicionamento teórico de Ohala, voltemos às diferentes maneiras de se considerar os fatos acústicos e articulatórios. Agora, trataremos da sua caracterização em um arcabouço dinâmico de produção e percepção da fala.

A partir da proposta da FAR (Browman & Goldstein, 1992), poderíamos conjecturar que o o sinal acústico não seria importante para o desenvolvimento da teoria. Do ponto de vista de um Programa de Pesquisa Científica de base dinâmica, a FAR tinha como objetivo tratar dos aspectos puramente articulatórios da produção da fala (ou da ação envolvida na sua produção – cf. Fowler & Goldstein, 2003). No entanto, ao considerarmos as propostas da TRD, conseguimos notar que o domínio acústico desempenha papel fundamental na efetivação do processo comunicativo, i.e., é possível notar que tanto o domínio acústico quanto o articulatório devem ser tomados como fundamentais para a produção e percepção da fala.

¹⁶⁹ “My own view is that between phonology and phonetics, phonology is the superordinate discipline, not because it has accomplished more or is better developed – the opposite may be true – but simply because it looks at and seeks answers to a much broader range of phenomena involving speech behavior. The answers to these questions will come, I believe, from phonetics, psychology, and studies of culture and society – including both the ways these domains determine speech behavior currently and in the past. Although in this paper I have emphasized to necessity of drawing on phonetics for phonological questions, I do not mean to neglect the resources in these other domains. [...] No doubt those who use the term 'interface between phonology and phonetics' believe that they are talking positive steps to break down the barrier between the two fields. But I maintain that if the term 'interface' preserves the outdated view that we are dealing with largely autonomous fields, it will impede our chances of improving on the past. I propose an alternative phrase and venture: the integration of phonology and phonetics” (Ohala, 1990: p.198).

Contrariamente às abordagens estritamente auditivas dos estruturalistas e simbólicas dos gerativistas, a dinâmica sugere que, embora o primitivo acional e perceptual sejam de base articulatória e real, o sinal acústico recebe uma posição de suma importância para a transmissão das informações acerca dos movimentos realizados pelos articuladores no interior do trato vocal. As ondas sonoras são fundamentais para que os estímulos sejam interceptados pelos usuários da língua a partir da estimulação dos órgãos do sentido que compõem o sistema perceptual humano.

Por exemplo, um indivíduo tem a tarefa de articular um determinado som da sua língua. Ele o faz através da imagem mental que tem estocada e que é relativa ao conjunto dos articuladores que têm de ser acionados para o cumprimento dessa tarefa. Aliado a isso, o falante sabe qual é o tempo de ativação de cada articulador, a maneira como cada articulador se coordena com os demais e o movimento que esse articulador tem de descrever (e.g., ele pode se encostar noutro ou apenas se aproximar maximamente dele). Toda essa complexa organização das tarefas envolvidas produz o sinal acústico que "carrega" ao receptor da mensagem todas essas informações que têm de ser interceptadas por ele, para que ele consiga "perceber" o que foi articulado.

Como se pode notar, acústico e articulatório são fundamentais para a percepção da fala no arcabouço da TRD. Ao contrário de se falar em decodificação ou tradução das informações articulatórias presentes no sinal acústico, é preciso tratar a percepção como um processo ativo de interceptação dos estímulos disponíveis nas mídias informacionais que carregam as características dos movimentos realizados pelos usuários das línguas. É essa característica direta e realista que possibilita a incorporação da variável tempo nas representações que, por sua vez, é percebida pelos percebedores de uma determinada língua.

Mais detalhadamente, a TRD propõe que o indivíduo que "percebe" o sinal acústico tem de depreender nele relações temporais entre os movimentos dos articuladores. Tais variações de relações temporais fazem com que se percebam, por exemplo, as diferenças entre um // de início de sílaba ou palavra como *light* e um de coda silábica ou de final de palavra como *dark*. Sproat & Fujimura (1993) não trabalharam com análises perceptuais sobre a natureza ora *light* e ora *dark* do // da

língua inglesa. A questão de seu trabalho não era se havia variantes *dark* ou *light*, e sim se só havia essas variantes. Se pensarmos categoricamente e se assumirmos a hipótese bigestual de organização das laterais do inglês é indiscutível que as diferentes relações temporais são veiculadas pelo sinal acústico e, conseqüentemente, percebidas. Para se corroborar essa hipótese ainda é preciso verificar se os usuários da língua percebem as diferenças gradientes que se distribuem no *continuum* entre uma categoria e outra.

Por fim, falta ainda mencionar que a proposta da TRD de dar importância de mesmo peso para os níveis acústico e articulatório só é viável se os envolvidos na percepção da fala forem tidos como usuários *ativos* da língua. Ao contrário da percepção *passiva* do estruturalismo e *parcialmente ativa* do gerativismo, o percebedor da TRD se desloca em seu nicho em busca de informações nas mais diferentes mídias (ar, tato ou luz refletida) para interceptar as fontes dos estímulos que alimentam os sistemas perceptuais. Ou seja, como os usuários são ativos, eles buscam outras fontes de estimulação na ausência de informações em uma das mídias. Essa é a base de um sistema multimodal de percepção que o faz totalmente distinto da aborgadem abstrata, simbólica e estática da TM. Um exemplo disso é quando tentamos conversar em um ambiente com um som muito alto. Diante dessa situação, imediatamente (e ativamente) buscamos informações visuais com relação aos movimentos dos articuladores de nosso interlocutor. Isso somente é possível se considerarmos a percepção como um sistema complexo que tem suas modalidades perceptuais ajustadas (e integradas) em decorrência das alterações (in)disponibilizadas pelo ambiente.

1.4 Sumário da seção: por que a polêmica não fecha?

Nas seções anteriores, apresentamos a polêmica em que se envolveram Ohala (1996) e Fowler (1996) sobre a natureza dos primitivos da percepção da fala. Nosso objetivo era de “ressuscitar” essa discussão para que tenhamos consciência de que ainda há muita pesquisa a ser feita antes de adotar um primitivo de análise perceptual ou outro.

Também tentamos mostrar no decorrer das seções deste capítulo os motivos pelos quais a polêmica não fecha. O fato de ela não fechar pode ser considerado um dado importante para a hipótese defendida neste trabalho, a saber: a escolha do primitivo da percepção é dada epistemologicamente, i.e., a depender da teoria fonológica escolhida é possível contemplar novos fenômenos que, por sua vez, abrem campo para a verificação experimental (a partir de um Programa de Investigação Científica) com relação à produção e percepção da fala.

No capítulo 2 deste trabalho, é possível verificar a partir de uma breve retrospectiva histórica (nada exaustiva) que cada teoria fonológica propõe novos primitivos de análise que acabam influenciando (ou guiando) as pesquisas sobre a percepção da fala. Tais estudos acabam por corroborar ou refutar conceitos sugeridos pelas teorias fonológicas. Um exemplo da influência dos pressupostos teóricos sobre a pesquisa em percepção da fala é a tentativa de os pesquisadores dos Laboratórios *Haskins* procurarem os invariantes acústicos que guiavam a percepção com o auxílio do *Pattern Playback*. Nesse mesmo exemplo, há o fracasso dessas tentativas enviesadas pelos princípios da Fonologia Estruturalista dos *Preliminaries* (Jakobson, Fant & Halle, 1952).

Diante do fracasso de se encontrarem os invariantes perceptuais no domínio acústico da fala, Liberman (1957) postula, ainda timidamente, as bases da Teoria Motora da Percepção da Fala (a *Early Motor Theory*). Sua proposta só ganha uma versão mais forte e robusta (em 1985) após os avanços propiciados pela Fonologia Gerativa Padrão (Chomsky & Halle, 1968), que possibilita um olhar mais abstrato e mentalista para a produção e percepção da fala. Da mesma forma, só foi possível propor a versão revisada da TM (Liberman & Mattingly, 1985) quando ocorrem os achados referentes ao *Efeito McGurk* (McDonald & McGurk, 1976) e a publicação da teoria modular da mente (Fodor, 1983).

Na “contramão” dos trabalhos de base gerativo-transformacional, há os trabalhos de Fowler (1981 e seguintes) sobre a influência de aspectos coarticulatórios na percepção da fala. É essa abordagem de base dinâmica que promove a proposição da TRD (Fowler, 1986). Essa teoria passa a trabalhar dentro de um arcabouço teórico de natureza distinta das anteriores. Abandona-se o olhar puramente simbólico (ou dicotomizado pela determinação do que é fonético ou

fonológico em uma dada língua), para que os aspectos reais envolvidos na produção e percepção da fala sejam considerados. É importante mencionar a relevância da proposição do gesto articulatório como primitivo de análise fonológica (Browman & Goldstein, 1986 e seguintes), pois é junto com essa abordagem teórica que a TRD dialoga e busca fenômenos passíveis de serem tratados em seu arcabouço teórico.

As seções que sucederam a apresentação da polêmica (no capítulo 3) mostraram que cada uma dessas teorias de percepção estavam vinculadas a programas de investigação científica que tratavam os componentes do fenômeno de perceber a fala de maneira distinta. Vale lembrarmos o olhar estritamente direcionado para os fenômenos acústicos do período estruturalista, que desconsiderava o receptor de uma mensagem como responsável pela sua decodificação.

Em contrapartida, a abordagem de base gerativo-transformacional da TM propõe um ouvinte parcialmente ativo que tem por objetivo converter a mensagem acústica em símbolos de base articulatória a partir de regras fonológicas que fazem essa tradução. Aqui, o sinal acústico estimula o sistema perceptual dos ouvintes para que se percebam as articulações pretendidas pelo falante da língua. Ou seja, o domínio acústico é um epifenômeno. O importante para essa teoria é a abstração articulatória pretendida no momento que se produz a fala.

Por sua vez, há o olhar da TRD que busca os invariantes articulatórios reais que são transmitidos pelo sinal acústico. Nessa perspectiva, os usuários da língua são considerados sujeitos ativos no processo de percepção da fala, pois os percebedores interceptam o sinal acústico em busca de informações acerca do seu ambiente de sobrevivência. Em outras palavras, tanto os aspectos acústicos quanto os articulatórios são importantes para a TRD. Ainda mais, por influência de *Os sentidos* de Gibson (1966), todos os sentidos passam a ser relevantes para a percepção dos seres. Trata-se de um olhar multimodal (e multissensorial) sobre a percepção, i.e., a percepção da fala é entendida de maneira não *especial*¹⁷⁰, mas

¹⁷⁰ O fato de a proposta da TRD não ser *especial* é uma oposição à abordagem da TM que previa mecanismos específicos (como o módulo fonético) para tratar da fala enquanto um “código especial”. Para aprofundar esta questão, consultem Fowler & Rosenblum (1991a) para críticas à especialidade e modularidade da fala, Fowler & Rosenblum (1991b) para problematizar o efeito duplex e Rosenblum (2005) para salientar os achados neurofisiológicos que embasam a proposta multimodal da percepção.

sim como uma capacidade ecológica que integra todos os sistemas perceptuais disponíveis nos seres (cf. Rosenblum, 2005).

Em suma, numa abordagem dinâmica, o gesto assume *status* central no modelo. O gesto passa a ser *também* o movimento real do articulador, não apenas o pretendido tal como queria a TM. Neste ponto deste trabalho, é possível sumarizar que a noção de gesto assume diferentes referentes no “mundo” ao se considerar as propostas de teorias fonológicas e de percepção da fala. Em outras palavras, o que era fonema para Jakobson, Fant & Halle (1952), não era para Chomsky & Halle (1968) e, muito menos, para Browman & Goldstein (1992).

Diante disso, salientamos, mais uma vez, que cada uma das teorias de percepção da fala apresentadas aqui está vinculada a uma teoria fonológica que acaba influenciando a escolha do primitivo perceptual. Se pensarmos estritamente na polêmica mencionada há pouco, podemos notar que Fowler trabalha dentro de uma perspectiva dinâmica de produção e percepção da fala (considerando a produção como um sistema matemático complexo e a percepção como multimodal), enquanto a TM, em uma estritamente simbólica. Por sua vez, Ohala (1996) apenas critica as abordagens articulatórias sem se vincular explicitamente a um viés teórico. Em sua argumentação é possível notar que as evidências apresentadas sugerem que há uma certa falta de embasamento experimental para tratar dos dados da percepção. Seria essa falha de natureza empírica que promoveria a implausibilidade de se adotar primitivos de natureza articulatória para a percepção. Talvez seja esse o ponto no qual os adeptos de uma perspectiva dinâmica sobre a linguagem estejam trabalhando. Faltam, contudo, diretrizes que orientem a coleta de dados que embasem empiricamente essa proposta.

Ao resgatarmos trabalhos anteriores de Ohala (e.g., Ohala, 1990), podemos conjecturar que o seu posicionamento contra os primitivos de base articulatória está relacionado ao fato de que seria necessário haver a tradução do nível acústico em um articulatório durante o processo de perceber a fala. Levando em conta, então, que para Ohala (1990) os níveis fonéticos e fonológicos são integrados, não haveria a necessidade de se traduzir um nível em outro. Portanto, a proposição de um primitivo de natureza articulatória seria desnecessária. A crítica de Ohala é claramente direcionada à abordagem gerativo-transformacional da TM, uma vez que

a da TRD não necessita de tradução de um nível em outro (cf. Fowler & Rosenblum, 1991). A TRD trata o fenômeno perceptual de maneira muito mais ampla. Não há apenas uma relação entre o acústico e o articulatório. De maneira geral, há uma relação entre o articulatório e as mídias de transmissão das características motoras, sejam elas dadas pelo ar, pela luz ou através do tato, ou seja, podemos dizer que se trata de uma abordagem multimodal e nada *especial* sobre a percepção. Tais questões sequer seriam de interesse para as outras abordagens sobre a percepção da fala.

CONSIDERAÇÕES FINAIS

Neste trabalho, tínhamos por objetivo apresentar um panorama reflexivo sobre as teorias de percepção da fala. Iniciamos essa breve história estabelecendo as diferentes maneiras de se conceber a natureza dos primitivos perceptuais adotados pelas teorias. No Capítulo 1, mostramos que há abordagens teóricas de base acústica e de base articulatória. Enquanto naquelas há, basicamente, proposições teóricas que tentavam averiguar a relação dos possíveis invariantes acústicos com a memória, a fonologia e a percepção da fala por outros animais, nestas há uma divisão das propostas, pois há uma de base mentalista e outra, realista.

A teoria de base mentalista aqui apresentada é a Teoria Motora da Percepção da Fala – TM – (Liberman & Mattingly, 1985). Nela, o primitivo perceptual é o *gesto pretendido* pelo falante. Essa proposta tenta dar conta da falta de relação biunívoca entre os invariantes acústicos e articulatórios descoberta na década de 1950 por Liberman e colegas (cf. Liberman, 1957) nos *Laboratórios Haskins*. Trata-se de uma proposta de base mental, pois ela sugere que os gestos pretendidos são símbolos disponíveis nos cérebros dos usuários de uma dada língua. Além disso, há a natureza estática e puramente simbólica dessa representação, aspecto este influenciado pela Fonologia Gerativa Padrão de *SPE* (Chomsky & Halle, 1968).

Por sua vez, a abordagem realista é encabeçada pelo trabalho de Fowler (Fowler, 1984 e seguintes) na sua Teoria do Realismo Direto da Percepção da Fala (TRD). Aqui, o primitivo é o *gesto articulatório*, que é uma unidade tanto simbólica quanto motora que é percebida diretamente pelos usuários da língua. Ou seja, para a TRD, os percebedores detectam as informações articulatórias que de fato foram realizadas por um dado falante. De acordo com a autora, isso é possível porque há invariantes articulatórios que são veiculados pelo meio acústico.

Tendo, então, apresentado as diferentes maneiras de se considerar a natureza dos primitivos da percepção da fala, traçamos no Capítulo 2 um breve histórico sobre como cada uma das abordagens teóricas se construiu. Salientamos as bases teóricas de cada uma delas vinculando-as aos cortes epistemológicos

realizados pelas teorias Estruturalista, Gerativista e Dinamicista.

Nesse capítulo, foi possível verificar que a percepção da fala somente podia ser considerada como audição dentro de um arcabouço estruturalista de estudos fonológicos. Tanto as ferramentas metodológicas para o tratamento do sinal da fala (seja por espectrógrafos ou fazendo uso do *Pattern Playback*) quanto as proposições teóricas para o tratamento fonêmico das línguas (tal como a disponível nos *Preliminaries*) possibilitavam apenas um olhar de natureza acústica para a percepção da fala.

Em contrapartida, na virada promovida pelos estudos linguísticos de base gerativo-transformacional, notamos uma alteração na maneira de se conceber os fenômenos linguísticos. Ou seja, lança-se um olhar verticalista sobre a linguagem, na qual uma mente dotada de capacidades inatas sobre a linguagem orienta a sua aprendizagem. Essa abordagem permite com que a TM seja elaborada, pois após a proposta da Fonologia Gerativa Padrão (do *SPE* – Chomsky & Halle, 1968) passa-se a tratar da produção e percepção da fala com um olhar mentalista sobre a linguagem. Essa nova maneira de considerar os fenômenos da linguagem ganha força com a descoberta do *Efeito McGurk* (McDonald & McGurk, 1976) e com a proposta modular da mente (Fodor, 1983). Esses marcos teóricos e empíricos fundamentam a adoção de um primitivo perceptual de base simbólica e estática, tal como a unidade disponível na versão revisada da TM (Liberman & Mattingly, 1985): o *gesto pretendido*. Em suma, as características assumidas pela unidade perceptual são as mesmas adotadas para a fonologia, a saber: são unidades simbólicas e estáticas que sequer precisam ser implementadas para serem percebidas.

Por fim, foi possível reconhecer que os avanços promovidos na coleta e análise de dados de fala iluminaram certos fenômenos que estavam à sombra de um olhar puramente estático e simbólico sobre a produção e a percepção. Os dados disponibilizados pela Fonologia Articulatória (Browman & Goldstein, 1986 e seguintes) e por Fowler (1981 e seguintes) sugeriam que o tratamento de base simbólico promovido pelo Gerativismo era insuficiente para acomodar a influência de aspectos coarticulatórios sobre a percepção da fala. É com esse panorama e sob a abordagem ecológica de Gibson (1966) que a TRD ganha sua primeira versão em 1986. O olhar dinâmico sobre a produção da fala possibilita que se trate a percepção

da mesma maneira, i.e., começa-se a considerar os aspectos multissensoriais envolvidos na percepção da fala e a considerar o usuário da língua como um sujeito ativo que busca em seu ambiente as informações relevantes para a sua sobrevivência. Em resumo, tal como a TM é fruto de marcos teóricos que dão força ao programa gerativista, a TRD se vincula a uma abordagem dinâmica e ecológica da linguagem. Tanto que o *gesto articulatório* passa a ser a moeda comum de troca (garantida pela *paridade* dos sistemas comunicativos) entre a produção e a percepção da fala.

Em outras palavras, tentamos mostrar no Capítulo 2 que a adoção dos primitivos perceptuais está vinculada aos cortes epistemológicos realizados pelas teorias fonológicas. Diante disso que no último capítulo, expusemos a polêmica de 1996 na qual se envolveram John Ohala e Carol Fowler. O objetivo de apresentar esse debate era de ilustrar, a partir de uma polêmica pública (cf. Dascal, 2006), que a escolha dos primitivos perceptuais está vinculada a arcabouços teóricos distintos. A não aceitação da natureza dos primitivos de percepção da fala pelas partes envolvidas na polêmica funciona como um exemplo de como o olhar sobre os fenômenos da fala é enviesado pelo programa de pesquisa que embasa cada uma das teorias de percepção.

Como pudemos ver, a polêmica não se desenvolve e tampouco fecha. Para nossa proposta, isso é um indício de que cada um dos debatedores trata a percepção da fala a partir de um ponto de vista específico. Enquanto Ohala trata a fala com a sua *visão particular* sem interfaces entre a fonética e a fonologia, Fowler encontra apoio em teorias de bases ecológica e dinâmica para propor a TRD. Como cada um se fecha em seu próprio loteamento¹⁷¹ epistemológico, a polêmica não se desenvolve.

Diante da nossa proposta, o leitor pode inferir que argumentamos contra a realização de debates, uma vez que cada uma das teorias se fundamenta sobre seu próprio terreno. Não! Acreditamos que os debates devem existir e, ao que parece, eles são fundamentais para que os proponentes de teorias revisem suas premissas e deixem-nas mais claras.

¹⁷¹ Empréstimo de Negri (2006) a metáfora de “loteamento de território” para a delimitação das áreas de atuação das disciplinas linguísticas.

No entanto, argumentamos que as críticas lançadas a uma dada teoria não podem fugir ao escopo de atuação de sua própria proposta. Por exemplo, tentar dar um tratamento gerativo-transformacional para os dados de alofonia gradiente é tão inviável quanto tentar buscar produções isoladas e totalmente livres de coarticulação dentro de uma abordagem dinâmica. Nesse sentido, estamos juntos com Borges Neto (2004) acerca da aceitação do *pluralismo teórico da linguística*, i.e., há tantas teorias quanto olhares distintos que se podem lançar sobre a linguagem. O nosso objetivo enquanto cientistas da linguagem é o de submeter as teorias a mais severa crítica obedecendo (e testando) os limites de cada proposta.

Em suma, ao invés de postularmos que as polêmicas “não levariam a nada” devido às respectivas especificidades epistemológicas, elas poderiam ser tomadas como pontos importantes para o direcionamento dos Programas de Investigação Científica, uma vez que, durante um debate, é preciso rever suas premissas teóricas e revisar quais são os fenômenos de interesse para o seu desenvolvimento. Em outras palavras, por mais que as polêmicas mantenham uma discussão “em aberto”, elas promovem o esclarecimento sobre os interesses da teoria. E seria sobre esses interesses específicos das teorias que as críticas devem ser fundamentadas.

Por fim, alguns pontos deste trabalho ainda merecem esclarecimento.

Primeiro, a maneira como traçamos este breve histórico acerca das teorias de percepção da fala pode sugerir uma preferência nossa sobre a abordagem dinâmica. Nosso critério de organização de apresentação das teorias se baseou simplesmente na cronologia das propostas teóricas. Nosso interesse era unicamente de descrever, entender e refletir como as teorias de percepção da fala foram construídas e sobre qual terreno foram levantadas. Ter consciência desses aspectos é fundamental para que o pesquisador saiba quais são os dados que são de interesse para uma determinada abordagem e como ele pode testar o limite do modelo adotado.

Segundo, quanto às cenas dos próximos capítulos, sugerimos a busca por outras polêmicas que também não se desenvolvam. Pode parecer “estranho” olhar para esses debates que não se desenvolvem, mas eles podem mostrar como as teorias lidam com os seus próprios limites. Ainda, olhar para essas polêmicas que “morrem” pode ampliar o entendimento sobre as disputas teóricas existentes na

linguística, pois há o aumento no número de discussões para análise. O que queremos dizer é que, ao contrário do que é de interesse para a Teoria das Controvérsias (Dascal, 1994), olhar para pequenas polêmicas pode indicar que cada abordagem trata de seus fenômenos de uma maneira distinta e isso impossibilita a comparação de teorias.

Por fim, salientamos que um olhar de natureza discursiva, que vincula a enunciação polêmica com a sua articulação ideológica, pode iluminar novas relações (tanto teóricas quanto relacionadas ao poder) que estão escondidas ao se adotar uma abordagem teórica ou outra (cf. Van Dijk, 2008). Mas isso é uma questão para desenvolvimentos futuros.

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, D. *A vida, o universo e tudo mais*. (O mochileiro das galáxias, v.3). Rio de Janeiro: Sextante, 2010.

ALBANO, E. C. *Da fala à linguagem: tocando de ouvido*. São Paulo: Martins Fontes, 1990.

ALBANO, E. C. *O gesto e suas bordas*. Campinas: Mercado de Letras/FAPESP, 2001.

ALBUQUERQUE, J. I. De A. *Aspectos da percepção da dessonorização terminal do inglês por falantes nativos de português brasileiro*. Dissertação de mestrado. Inédito. Curitiba, UFPR: 2012.

BECKMAN, M. & KINGSTON, J. (eds.). *Papers in Laboratory Phonology*. Cambridge University Press, 1990.

BINNIE, C.A.; MONTGOMERY, A. A.; JACKMAN, P. L. Auditory and visual contributions to the perception of consonants In: *Journal of Speech Hearing Research*, 17, 619-630, 1974.

BORGES NETO, José. *Ensaio de Filosofia da Linguística*. São Paulo: Parábola, 2004.

BROSNAHAN, L. F.; MALMBERG, B. *Introduction to phonetics*. Cambridge: Heffer, 1970.

BROWMAN, C. & GOLDSTEIN, L. Articulatory gestures as phonological units. In: *Phonology*, 6, 201-251. 1989

BROWMAN, C. & GOLDSTEIN, L. Tiers in articulatory phonology, with some implications for casual speech. In: KINGSTON, T.; BECKMAN, M.E. (Eds.). *Papers in Laboratory Phonology I: between grammar and physics of speech*. pp: 341-376. Cambridge University Press, 1990.

BROWMAN, C. & GOLDSTEIN, L. Towards an articulatory phonology. In: *Phonology Yearbook*, 3, 219-252. 1986

BROWMAN, C.; GOLDSTEIN, L. Articulatory Phonology: an overview. In:

Phonetica, 49, pp. 155-180, 1992.

CÂMARA JR., J. M. *Estrutura da Língua Portuguesa*. 2. ed. Petrópolis/Rio de Janeiro: Vozes, 1970.

CANDIOTTO, K. B. B. Fundamentos epistemológicos da teoria modular da mente de Jerry A. Fodor. In: *Trans/Form/Ação*. São Paulo, 2008. pp. 119-135.

CARNEY, A. E.; WIDIN, G. P.; VIEMEISTER, N. F. Non categorical perception of stop consonants differing in VOT. In: *Journal of Acoustical Society of America*, 62. 1977. pp. 961-970.

CHOMSKY, N. & HALLE, M. *The Sound Pattern of English*. New York: Harper & Row. 1968.

CHOMSKY, N. A review of B. F. Skinner's Verbal Behavior. In: *Language*, 35(1). 1959. pp.26-58.

CHOMSKY, N. *Aspects of the Theory of Syntax*. Cambridge, Mass: The MIT Press, 1965. (Trad. portuguesa, Coimbra: Arménio Amado, 1978)

CLEMENTE, F. C. *Retroflexão gradiente nos róticos em coda do PB de Curitiba*. Dissertação de Mestrado. Inédita. Programa de Pós-Graduação em Letras, UFPR: 2009.

COLE, R. A. & SCOTT, B. Toward a theory of speech perception. In: *Psychological Review*, 81. 1974. pp. 348-374.

CROWDER, R. G. & MORTON, J. Pre-categorical acoustic storage (PAS). In: *Perception and Psychophysics*, 5. 1969. pp. 365-373.

D'ANGELIS, W. da R. *Traços de modo e modos de traçar geometrias: línguas Macro-Jê & teoria fonológica*. Tese de doutorado. 2 volumes. Campinas: IEL/UNICAMP, 1998.

DASCAL, M. & BORGES NETO, J. De que trata a linguística, afinal? In: BORGES NETO, J. *Ensaio de filosofia da linguística*. São Paulo: Parábola Editorial, 2004.

DASCAL, M. *Interpretação e compreensão*. São Leopoldo: Editora da Unisinos, 2006.

DASCAL, Marcelo. "Epistemologia, Controvérsias e Pragmática". In: *Revista da Sociedade Brasileira de História da Ciência (SBHC)*, n. 12, 1994, p. 73-98.

DELATTRE, P. C.; LIBERMAN, A. M., & COOPER, F. S. Acoustical loci and transitional cues for consonants. In: *Journal of the Acoustical Society of America*, 27, 769-773, 1955.

DIEHL, R.; LOTTO, A.; HOLT, L. Speech perception. In: *Annual review of Psychology*, vol 55. 2004. pp. 149-179.

FANT, G. *Acoustic theory of speech production*. Haia: Mouton, 1960.

FÁVERI, R. B. *Fundamentos para uma filosofia da linguística: a controvérsia sobre a natureza do componente semântico e o conceito de estrutura profunda em gramática gerativa transformacional*. Tese de doutorado. Inédito. Programa de Pós Graduação em Letras. Curitiba: UFPR, 2011.

FEYERABEND, P. Consolando o especialista. In: LAKATOS & MUSGRAVE (eds). *A crítica e o desenvolvimento do conhecimento*. São Paulo: Cultrix, 1979.

FODOR, J, A. *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press , 1983.

FOWLER, C. A. & SALTZMAN, E. Coordination and Coarticulation in Speech Production. In: *Language and Speech*, 36. 1993. pp.171-195.

FOWLER, C. A. Production and perception of coarticulation among stressed and unstressed vowels. In: *Journal of Speech and Hearing Research*, 46, 1981. pp. 127-139.

FOWLER, C. A. Segmentation of coarticulated speech in perception. In: *Perception & Psychophysics*, 36, 1984. pp. 359-368.

FOWLER, C. A. "Perceptual centers" in Speech production and perception. In: *Perception & Psychophysics*, 25, 1979. pp.375-388.

FOWLER, C. An event approach to the study of speech perception. In: *Journal of Phonetics*, 14, pp. 3-28, 1986.

FOWLER, C. Coarticulation and theories of extrinsic timing control. In: *Journal of Phonetics*, 8, 1980. p.113-133.

FOWLER, C. Listeners do hear sounds, not tongues. In: *The Journal of the Acoustical Society of America*, 99 (3), pp. 1730-1741, 1996.

FOWLER, C. Phonological and Articulatory Characteristics of Spoken Language. In: BLANKEN, G.; DITTMANN, J.; GRIMM, H.; MARSHALL, J. C.; WALLECH, C- W. (eds.) *Linguistic Disorders and Pathologies: An International Handbook*. Berlin, New York: Walter de Gruyter, 1993. pp.34-46.

FOWLER, C.A. & ROSENBLUM, L.D. Duplex perception: A comparison monosyllables and slamming doors. In: *Journal of Experimental Psychology: Human Perception and Performance*. 16, 1991b. pp.742-754.

FOWLER, C.A. & ROSENBLUM, L.D. Perception of the phonetic gesture. IN: I.G. MATTINGLY, I. G. & STUDDERT-KENNEDY, M. (Eds.). *Modularity and the Motor Theory*. Hillsdale: Lawrence Earlbaum, 1991a. pp. 33-59.

FOWLER, C. & DEKLE, D.J. Listening with eye and hand: crossmodal contributions to speech perception. In: *J. Exp. Psychol. Hum. Percept. Performance* 17, 816-828, 1991.

FREITAS, M. C. C. de. *O gesto fônico na aquisição desviante: movimentos entre a produção e a percepção*. Tese de doutorado. Inédita. Campinas: Unicamp/IEL, 2012.

GIBSON, J. J. *The ecological approach to visual perception*. Boston: Houghton Mifflin, 1979

GIBSON, J. J. *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.

GOLDSTEIN, L. & FOWLER, C. A. Articulatory phonology: A phonology for public language use. In: Schiller, N.O. & Meyer, A.S. (eds.), *Phonetics and Phonology in Language Comprehension and Production*, Mouton de Gruyter, 2003. pp. 159-207.

HAMMARBERG, R. The metaphysics of coarticulation. In: *Journal of Phonetics*, 4. 1976. pp.353-363.

HOCKETT, C. *Manual of phonetics*. Bloomington: Indiana University, 1955.

ILARI, R. (2005). O estruturalismo linguístico: alguns caminhos. In: MUSSALIN & BENTES (Org). *Introdução à linguística: fundamentos epistemológicos*. Volume 3. 2ª edição. São Paulo: Cortez, 2005.

JAKOBSON, R.; FANT, G.; HALLE, M. *Preliminaries to speech analysis*.

Cambridge: MIT Press. 1952.

KEATING, P. Universal phonetics and the organization of grammars. In: FROMKIN, V. (Ed.). *Phonetic Linguistics: essays in honor of Peter Ladefoged*. New York: Academic Press, 1985. p. 115-132.

KUHL, P. K. & IVERSON, P. Linguistic experience and the “perceptual magnet effect”. In: STRANGE, W. (Ed.). *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press, 1995.

KUHL, P. K. & MILLER, J. D. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. In: *Science*, 190. 1975.

KUHL, P. K. Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. In: *Journal of Phonetics*, 21, 125-139, 1993.

KUHL, P. K. Human adults and human infants show a “perceptual magnet effect” for prototypes of speech categories, monkeys do not. In: *Perception & Psychophysics*, 50, 93-107, 1991.

KUHN, T. S. *A Estrutura das Revoluções Científicas*. São Paulo: Perspectiva, 1975[1962].

KUHN, T. S. *O caminho desde a estrutura*. São Paulo: Editora da UNESP, 2006.

LADEFOGED, P. & JOHNSON, K. *A course in phonetics*. 6.ed. Boston: Wadsworth/Cengage Learning, 2011.

LADEFOGED, P. *A course in phonetics*. 4. ed. Boston: Heinle & Heinle, 2001.

LADEFOGED, P. On dividing phonetics and phonology: comments on the papers by Clements and by Browman and Goldstein. In: BECKMAN, M. & KINGSTON, J. (eds.). *Papers in Laboratory Phonology*. Cambridge University Press, 1990. pp. 398-405.

LAKATOS, I. O falseamento e a metodologia dos programas de pesquisa científica. In: LAKATOS & MUSGRAVE (eds). *A crítica e o desenvolvimento do conhecimento*. São Paulo: Cultrix, 1979.

LAMEIRA, A. P; GAWRYSZEWSKI, L. de G.; PEREIRA JR., A. Neurônios

-
- espelho. In: *Psicologia USP* [online]. 2006, vol.17, n.4, pp. 123-133.
- LIBERMAN, A. M. & STUDDERT-KENNEDY, M. Phonetic perception. In: HELD, R.; LEBOWITZ, H. W.; TEUBER, H. L. (eds). *Handbook of Sensory Physiology*, vol. VIII: Perception. New York: Springer-Verlag, 1987.
- LIBERMAN, A. M. & WHALEN, D. "On the relation of speech to language." In: *Trends in Cognitive Sciences*, 4. 2000. pp.187-196
- LIBERMAN, A. M. *Speech: a special code*. Cambridge: MIT Press, 1996.
- LIBERMAN, A. M., ISENBERG, D., & RAKERD, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. In: *Perception & Psychophysics*, 30, 133-143.
- LIBERMAN, A. M.; COOPER, F. S.; SHANKWEILER, D. P.; STUDDERT-KENNEDY, M. Perception of the speech code. In: *Psychological Review*, 74. 1967. pp. 431-461.
- LIBERMAN, A. M.; DELATTRE, P. C.; COOPER, F. S. Acoustic loci and transitional cues for consonants. In: *Journal of the Acoustical Society of America*, 27. 1955. pp. 769-773.
- LIBERMAN, A. M.; DELATTRE, P. C.; COOPER, F. S. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. In: *American Journal of Psychology*, 65. 1952. pp. 497-516.
- LIBERMAN, A. M.; MATTINGLY, I. G. The Motor Theory of Speech Perception Revised. In: *Cognition*, 21, pp.1-36, 1985.
- LIBERMAN, Alvin. (1957). Some results of research on speech perception. In: LIBERMAN, A. *Speech: a special code*. Cambridge, MA: MIT Press, 1996.
- LOTTO, A. J.; KLUENDER, K. R.; HOLT, L. L. Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). In: *Journal of the Acoustical Society of America*, 102. 1997. pp. 1134-1140.
- MALMBERG, Bertil (ed). *Manual of Phonetics*. Amsterdam: North-Holland Publishing Co, 1968.
- MATTINGLY, I. G., & LIBERMAN, A. M. Verticality unparalleled. Commentary/Fodor: Modularity of mind. In: *The Behavioral and Brain Sciences*, 8. 1985: 1, 24-26.

MCGURK, H. & MACDONALD, J. Hearing Lips and Seeing Voices. In: *Nature*. 1976. p. 264746-48.

MCKAIN, K.; STUDDERT-KENNEDY, M.; SPIEKER, S.; STERN, D. Infant intermodal speech perception is a left-hemisphere function. In: *Science*, 219, 1347-1349, 1983.

MUSSALIN & BENTES (Org). *Introdução à linguística: fundamentos epistemológicos*. Volume 3. 2ª edição. São Paulo: Cortez, 2005.

NEGRI, L. *Zona de fronteira: a delimitação entre a semântica e a pragmática sob a lente das expressões de polaridade negativa*. Tese de doutorado. Inédita. Curitiba: UFPR, 2006.

NISHIDA, G. Estudo piloto sobre a percepção do tap em grupos do PB. In: *Eletras*, vol. 23, n.23, dez. 2011. Disponível em: http://www.utp.br/eletras/dossie/artigo/Dossie_especial_artigo_23.3_Estudo_piloto_sobre_a_percepcao_do_tap_em_grupos_do_PB_Gustavo_Nishid.pdf.

O'CALLAGHAN, C. & NUDDS, M. Introduction: The philosophy of sounds and auditory. In: O'CALLAGHAN, C. & NUDDS, M. (orgs). *Sounds and Perception*. Oxford University Press, 2009. pp. 1-25.

ODEN, G. C. & MASSARO, D. W. Integration of featural information in speech perception. In: *Psychological Review*, 85. 1978. pp.172-191.

OHALA, J. J. The origin of sound patterns in vocal tract constraints. In: MACNEILAGE, P. F. (ed). *The Production of Speech*. New York: Springer-Verlag, 1983. pp. 189–216.

OHALA, J. Speech perception is hearing sounds, not tongues. In: *Journal of the Acoustical Society of America*, 99.1996. pp.1718-1725.

OHALA, J. There is no interface between phonology and phonetics. In: *Journal of Phonetics*, 18. 1990. pp.153-171.

ÖHMAN, S. Coarticulation in VCV utterances: spectrographic measurements. In: *Journal of Acoustical Society of America*, 39: 151-168, 1966.

PISONI, D. B. & TASH, J. Reaction times to comparisons within and across phonetic categories. In: *Perception & Psychophysics*, 15. 1974. pp. 285-290.

PISONI, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. In: *Perception and Psychophysics*, 13, 1973. pp.253-260.

PISONI, D. B. Identification and discrimination of the relative onset of two component tones: implications for the perception of voicing in stops. In: *Journal of Acoustical Society of America*, 61. 1977. pp.1352-1361.

RIZZOLATTI, G.; FADIGA, L.; GALLESE, V.; & FOGASSI, L. Premotor cortex and the recognition of motor actions. In: *Cognitive Brain Research*, 3, 1996. pp.131-141.

Rizzolatti, G.; Fadiga, L.; Gallese, V.; Fogassi, L. Premotor cortex and the recognition of motor actions. In: *Cogn. Brain Res.*, 3, 131-141, 1996.

ROSENBLUM, L. D. The primacy of multimodal speech perception. In: PISONI, D. & REMEZ, R. (Eds). *The Handbook of Speech Perception*. Malden: Blackwell, 2005. pp. 51-78.

SAGEY, E. *The Representation of Features and Relations in Nonlinear Phonology*. Cambridge: MIT dissertation, 1986.

SALTZMAN, E. Task dynamic coordination of the speech articulators. In: HEUER, H& FROMM, C. (Eds.), *Generation and modulation of action patterns*. (Experimental Brain Research Series 15). New York: Springer-Verlag, 1986. pp. 129-144.

SALTZMAN, E., & KELSO, J. A. S. Skilled actions: A task dynamic approach. In: *Haskins Laboratories Status Report on Speech Research*, SR-76,3-50. 1987.

SAUSSURE, Ferdinand de. *Curso de linguística geral*. (Org. Charles Balley e Albert Sechehaye). São Paulo: Cultrix, 1969.

SIILVA, A. H. P. *As fronteiras entre Fonética e Fonologia e a alofonia dos róticos iniciais em PB: dados de dois informantes do sul do país*. Tese de Doutorado. Inédita. Universidade de Campinas/IEL/LAFAPE, 2002.

SPROAT, R. & FUJIMURA, O. Allophonic variation in English /l/ and its implications for phonetic implementation. In: *Journal of Phonetics*, n. 21. 1993. pp. 291-311.

STERIADE, D. Gestures and autosegments. In: KINGSTON, T.; BECKMAN, M.E (Eds.). *Papers in Laboratory Phonology I: between grammar and physics of speech*. pp: 382-397. Cambridge University Press, 1990.

STEVENS, K.N. & BLUMSTEIN, S.E. Invariant cues for place of articulation. In: *J.*

Acoustic Soc. A m. 64(5), Nov. 1978

STUDDERT-KENNEDY & GOLDSTEIN. Launching language: The gestural origin of discrete infinity. In: CHRISTIANSEN & KYRBY (eds). *Language Evolution*. Oxford: Oxford University Press, 2003. pp. 235-254.

STUDDERT-KENNEDY, M. Speech perception. In: LASS, N. J. (ed.) *Contemporary issues in experimental phonetics*. New York: Academic Press, 1976. pp. 243-293.

TRASK, R. L. *Dicionário de Linguagem e linguística*. Tradução de Rodolfo Ilari. São Paulo: Contexto, 2004.

TRUBETZKOY, N. *Principes de Phonologie* [Grundzuge der Phonologie]. Paris, 1970 [1939].

VAN DIJK, T. A. *Discurso e poder*. Organização de Judith Hoffnagel e Karina Falcone. São Paulo: Contexto, 2008.

WILLIAMS, M. *Problems of Knowledge: A Critical Introduction to Epistemology*. Oxford University Press, 2001.